

British university crisis (cont'd)

The next phase in the depressing history of British universities is about to open. The only hope is to make the best of what lies in store.

THOSE who live in crisis are not best placed to tell when crisis turns to disaster. That is the present condition of the British universities, which have been worrying about their survival for the past five years, forever hoping for an end to their empty agony — and forever being disappointed. That time is plainly no nearer. During the past week, reports have been circulating that the University Grants Committee (UGC), in its role as the British government's adviser on university spending, is about to tell the Department of Education and Science that the time has come to decide which universities should no longer be supported from public funds in the usual way. What seems to have happened is that UGC, anticipating the detailed advice it will next month be giving the universities and the government about university budgets during the four years beginning in August, expects that, by the end of that period, some budgets will have shrunk to the point at which the institutions concerned will no longer be viable. Of necessity, the figures are not yet final, but the general drift of the arithmetic is not a surprise. UGC has been saying for the past eighteen months that the cash limits within which British universities will have to manage their affairs imply a continuing reduction of the resources available at the rate of two per cent a year. The logic of that position has been plain: sooner or later, the funds would not suffice to keep in being all the institutions now on the public payroll. The question has been not whether universities would have to be shut, or otherwise put out into the cold, but when — and which. UGC, recognizing that the choice of particular universities would become intensely political, has consistently said that the government will have to take responsibility for this invidious choice.

Budgets

There are a few months before that fur will begin to fly, although they will bring very little peace of mind to British academics. Next month, UGC will be writing to universities with some indication of how funds will be shifted from some universities to others to reflect its judgement of their relative strength in research and scholarship. On this occasion, the intention will be to concentrate resources on the institutions that are already strong. Those deemed to be at the other end of the spectrum will know which way the wind will be blowing when UGC later tells the government more formally about the implications of its decisions. Members of Parliament in the localities affected will be showered with invitations and will repay the hospitality with what they have to say in the House of Commons. That, at least, is what will happen unless the government changes its tune about its willingness to support the British university system on something like the scale on which it has been willed into existence by previous governments. The real surprise in the past few weeks is that there are signs that the hard line of the past five years may be relaxed, not because of concern that resources may otherwise be wasted but from a calculation that it would be imprudent to be battling with the universities as well as the schools when a general election is looming.

So what is the outcome most to be looked for in these changing circumstances? The temptation to hope that the government will find some extra money and that the British university system

will then be able to continue as it is at present for a few years more should be resisted. The changes that will become apparent next month will be welcome steps towards the moulding of a university system that embodies the kind of diversity that Britain has conspicuously lacked in the decades since the Second World War. For too long, the pressures of the UGC system have encouraged a uniformity of goals among British universities that has been both unrealistic and undesirable. Institutions that might have wished, or have been well-advised, to become liberal arts colleges have been discouraged by the models used for assessing performance. Even those that might have nursed ambitions to contribute more directly to faltering British technology than by training students have been seduced by the Oxbridge model of success into debilitating academic competition. Only with the beginning of the present phase of adversity have some institutions successfully broken away from the mould. The best example is the University of Salford, whose budget was cut by 41 per cent over a period of three years and which responded by turning itself a more practically-minded institution than it would have otherwise have become; that shows that sufficiently energetic institutions can successfully survive even apparently fatal budget cuts; but it is also true that in the British climate, not all those setting out on the Salford path have been successful.

Responsibility

The best hope now is that the government will use such conscience money as it can scrape together to help the institutions at the thin end of UGC's budget proposals to find some useful role as institutions of higher education in some other image than that of the conventional university. At the same time, it is essential that all universities on the public payroll should be given a greater sense of responsibility for the resources at their disposal than they at present enjoy, for only in that way can the system as a whole respond to the market forces with which it will be increasingly concerned in the years ahead. If some institutions threatened with closure should decide to make their future as post-experience institutions, as management schools or institutions the like of which has not yet been seen, and if they should thereby satisfy some real demand for higher education, neither the government nor UGC should demur. But help will be needed to make the transition. To fail to provide it will be further to waste intellectual resources.

This is also (looking on the bright side) the occasion when something should at last be done to bridge the present gulf between the universities and the rest of British higher education. For twenty years, since the casual invention of what is called the binary system of higher education, there have been two tiers of institutions that allow young people to earn degrees. But the polytechnics, which were devised so as to offer young people an opportunity to follow professional courses of study, have often had a paradoxical hankering after what they regard as university status. As it happens, in the past few weeks the government's coordinating body for the polytechnics has stirred the pot by publicizing its plans for living on the short commons it expects when the pressures soon to afflict UGC begin to press on what is



called the public sector. The government has responded indignantly, by saying that this other dismal future has not yet been decreed. But the coincidence of apprehension among the universities and the polytechnics should be an opportunity for the government to weld the two sectors together.

But will there be the students to fill all these institutions? Over the past five years, the British government has fought a running battle of words with British academic institutions over the likely demand for higher education in the years ahead. Government officials have over-anticipated the point at which the demographic retreat of the 1960s would make itself apparent in reduced demand for higher education, and have consistently been beaten in the argument. Now the official word is that the decline of demand will not arrive until 1990, still just beyond the planning horizon. But there is a sense in which this argument is irrelevant and arcane. The participation rate in British higher education is too low for a state with modern pretensions, while there is an urgent need that Britain should follow most other modern states in providing a four-year education for a greater proportion of its young people, partly for its own sake but crucially so as to allow the cramping specialization of the British high-school curriculum to be decoupled from young people's fateful choice of careers. And then there is the unmet (and, by taxation, repressed) demand for post-experience education.

Where will all this lead? Nowhere, if everybody does nothing. Almost nowhere if only some are prepared to recognize the present opportunity for what it is. But there is just a chance that the coincidence of the latest threat of crisis in British higher education and the prospect that there is electoral advantage to be won will create a sufficient sense of the importance of the occasion to suggest that the time has come radically to modernize an outdated system. That is the best and the only hope. □

Insect pests rampant

Techniques for managing resistance to pesticides have become essential in the United States.

CHARLES Darwin would readily have understood why the proportion of US crops lost to insects has increased from about 7 per cent in the 1940s to about 13 per cent today. The underlying cause of this alarming development is the repeated use of insecticides (of which the vast majority fall into just four basic chemical types), which has led by means of simple natural selection to the evolution of resistant strains of many of the most important pests. Commercial insect pests have now evolved resistance to all major classes of insecticides, including the most recent addition to the insecticide armoury, the synthetic pyrethroids, and the problem is undoubtedly going to get worse before it gets better. The usual response of the farmer when a resistant strain emerges is to switch to another pesticide, which is fine (although it may be expensive) as long as there are other materials still available. But the cost to the United States of the extra pesticide treatments made necessary by resistance, together with related crop losses, has been estimated to be at least \$150 million a year and could be much higher. There has been at least one instance where an insect pest became resistant to all applicable pesticides — which put an end to cotton-growing in much of northern Mexico. And the problem is not restricted to insects; other types of pests have also evolved resistance to the principal components of the chemical armoury.

Yet now there is some reason to fear that the renewal of the chemical armoury may have become more difficult. There are indications that the high cost of registering a pesticide with the Environmental Protection Agency (EPA) together with the poor outlook for product lifetime is deterring some manufacturers from developing new products. To be sure, manufacturers have made some efforts to provide farmers with fertilizer schedules that delay the emergence of resistance; it is after all in their interest as well as that of the farmers that products devel-

oped and nursed through expensive federal safety testing should retain their usefulness (and profitability) as long as possible. But understanding of the evolution of resistance, despite its economic importance, is still insufficient to allow anything more precise than educated guesses about the consequences of different strategies for the management of resistance.

This is one reason why there is an urgent need for more research. As existing products are withdrawn from the market because of safety fears or resistance, there will be an urgent need for chemicals aimed at new biochemical targets that avoid the dangers posed by cross-resistance to related agents. Who will pay? Industry is wary of such a basic research approach, for understandable reasons: it is inherently risky. Far safer simply to screen and optimize for the short-term biological effectiveness of derivatives of useful known compounds. The US Department of Agriculture, however, does have within its mandate the conduct of fundamental research that could lead to new pesticides, but so far has failed to do very much that is worthwhile. Instead, by means of their responses to research grant proposals, the National Science Foundation and the National Institutes of Health have shouldered most of the burden. The Department of Agriculture should take notice that this is an area of research that has clear potential for long-term economic benefit, but in which there is much basic science still to be learned. If ever there was a case for federal funding of basic research, this is it.

The regulators also need to be more diligent. The US Environmental Protection Agency (EPA) has formal responsibility for the safety and efficacy of pesticides, and has the authority to consider likely problems such as resistance when making licensing decisions. A recent study by the US National Research Council has come down against any formal extension of EPA's regulatory authority (*Pesticide Resistance: Strategies and Tactics for Management*) but at the same time has revealed how much more there is that EPA could be doing now to encourage responsible pesticide use. During President Reagan's first term of office, the troubles at EPA caused by the politicking of administrator Anne Burford left it with little time to worry about the environment outside Washington, and even today the division of EPA responsible for strategies for resistance — the integrated pest management division — is hamstrung by a non-existent extramural research budget that limits EPA's activities to those for which financial support can be found elsewhere. But even if EPA had the resources to do a credible job of collating information on pesticide resistance and issuing appropriate advice, the bulk of the much-needed applied research into the development of resistance (as opposed to basic research into biochemical mechanisms) would fall on industry. This is as it should be; industry will be the clear benefactor, and is better placed to do the job than government. The effort might even yield pesticides free from the problem of resistance.

At the same time, industry is right to be fearful of Congress giving EPA more statutory responsibility to manage pesticide resistance, at least until the power it already has is used effectively. All relevant experience supports the idea that local and informal decision-making, together with voluntary codes of conduct, are more likely to produce effective management schemes than the heavy hand of statute. One obvious need is that there should be coordinating recommendations on use of different chemicals over long time-periods. Antitrust laws that make companies shy of formulating anti-resistance plans with their competitors are only one example of the practical obstacles that have limited resistance management. While "integrated pest management" schemes introduced over the past decade have undoubtedly helped to delay some instances of resistance, it is equally clear that the scale and nature of the problem requires that industry should be given extra encouragement to develop more effective programmes. Cooperation with state and federal agencies is also needed. In the long run, this may be the only way in which to ensure that better products are eventually produced, and that better use is made of them. □

Yellow rain

Canada publishes more equivocal evidence

Washington

CANADIAN studies of yellow rain samples published last week by the Canadian Defense Research Establishment are being claimed as support both by those who believe the phenomenon to be due to chemical warfare and those who believe it to be merely mass defecation by bees. The study* finds small amounts of tricothecene toxins in some of the samples collected by Canadian military medical teams from Thailand in 1982, but notes that the levels are similar to those found in stored cereals infected with *Fusarium* fungus and concludes that "it is difficult to say if there is any chemical warfare significance".

Supporters of the US government's claim that tricothecene toxins have been used by Vietnamese and Laotian forces against civilians are, however, more impressed by higher levels of tricothecenes found on a plastic bag handed to the Canadian team by a villager from Ban Sa Tong which they believe could be part of a toxic weapon.

In the Ban Sa Tong incident, which occurred on 19 February 1982, villagers reported that an aircraft circled the village at 5,000 feet before releasing yellow rain. As most of it fell on just six houses, supporters of the weapon theory have had to sup-



pose that the toxin was dispersed at a much lower level. Evidence of a dispersion device has so far been lacking, yet William Kucewicz of the *Wall Street Journal's* editorial board wrote on 31 March that the Canadian report provided "the most conclusive proof yet that yellow rain is a man-made weapon" and that it "rules out the possibility these particular mycotoxins were naturally occurring".

Kucewicz's article, which appeared be-

fore the report was published, has prompted the Canadian government to express concern about misrepresentation of the report to the US government — the apparent source of the "essential details" leaked to Kucewicz. The Canadian Ambassador has written to the *Journal* and to a Thai newspaper that reprinted Kucewicz's article saying the report "does not reach conclusions" on possible use of toxic weapons.

All the plant samples examined by the Canadians proved to consist mainly of pollen, consistent with the hypothesis of Matthew Meselson of Harvard University that the characteristic yellow spots found on leaves result from mass defecations by bees, a well-recognized phenomenon. Where tricothecenes were present, they were generally only just distinguishable from background noise by gas chromatography and mass spectrometry. The report notes that, since little is known about the natural occurrence of tricothecenes in South-East Asia, comparisons with natural levels and types are "difficult if not impossible". None of the biomedical samples from exposed villagers discussed in the report showed evidence of tricothecenes, and there were no deaths caused, although an earlier Canadian study found an increase in cold-like symptoms.

The bag, in contrast — torn and dusted with a brown powder — did not contain pollen, although the Canadians did not apparently analyse the dust completely. The report is properly circumspect about the provenance of the bag, which unlike other samples was not collected by the Canadians at first hand but was offered by a villager some weeks after the 19 February incident, during a follow-up visit. Meselson argues that a bag used as part of a delivery system would have been found sooner during the intensive investigations and that it might be expected to show evidence of burn marks from explosives; whatever was on the bag could not be the same as "yellow rain" because of the absence of pollen. The levels of toxin found on the bag (one analysis showed 230 p.p.m. of T-2 toxin in the dust, although a later check found only 6 p.p.m.) are not inconsistent with a natural origin; in the laboratory, 900 p.p.m. of T-2 have been obtained in barley infected with *Fusarium*.

Meselson says the Canadians have presented the best-ever documented case of bee defecation but complains that the "rather pathetic analysis" of the plastic bag fails to provide sufficient information to know what absolute quantities of toxins were present. Meselson is sticking with bees for the later "yellow rain" reports, but does not exclude the possibility of some other chemical weapons having been used in Cambodia during the late 1970s.

Tim Beardsley

*Final Summary Report on the Investigation of "Yellow Rain" Samples from South-East Asia.

Yellow rain

British assays negative

THE *Wall Street Journal* article that has incensed the Canadians also referred to measurements made by British laboratories, which are said to have confirmed the presence of tricothecenes in samples of vegetation from South-East Asia, provided from the United States. But even though the results of the British measurements, carried out at the Chemical Defence Research Establishment of the British Ministry of Defence, are known to have been negative and not positive, the British authorities have not yet reacted with matching indignation at the *Journal's* misrepresentation.

The British experiments were carried out with great care, after the difficulties of the analysis of natural samples for quantities of tricothecenes of the order of parts per billion had been fully assessed. The work occupied a period of several months during 1983, and cost some £100,000. The negative results have been reported to the US defence laboratories involved in the analysis. There was an opportunity for a formal comparison of the results obtained at half a dozen laboratories, mostly in the United States, at a private symposium last year after a Gordon conference on tricothecenes held in New Hampshire.

Since the significance of the British measurements became apparent to the

British authorities, there has been a careful evaluation of the manner in which they were carried out. This, however, has not shaken the confidence of those responsible for the measurements or of their managers in the general conclusion that the samples from South-East Asia embody no decisive evidence that yellow rain is linked with artificially engendered tricothecene toxins.

The *Wall Street Journal* has had an extreme position on the supposed Soviet use of chemical weapons for at least the past two years. In 1983, William Kucewicz published a series of articles in the *Journal* which, among other things, alleged that Soviet laboratories were using genetic manipulation to manufacture influenza viruses that could be used as biological warfare agents. Last year, he published a fierce attack on Professor Matthew Meselson of Harvard University after the latter had published an account of his bee faeces theory in *Scientific American*.

A formal report of the British measurements is believed to have been circulated among British government departments. A spokesman for the Ministry of Defence said earlier this week that investigations were continuing, and that it was the present intention to publish the results.

John Maddox

French science

Living with budget cuts

FRENCH research may be facing a budget cut, if the new cost-conscious government of right wing prime minister Jacques Chirac has its way. Late last week, the exact size of the cut had not been officially announced, but sums of the order of FF1,500 million (£150 million), or around five per cent of the government civil research and development budget, were being suggested.

Research minister Alain Devaquet was attempting to muster support in the council of ministers for a less damaging attack on his bailiwick, which the powerful minister of state for the economy, finance and privatization, Edouard Balladur, and his colleague for the budget, Alain Juppé, appear to consider a prime target for government savings.

But other fears in the research community seem to be receding. In particular, the prospect that the ministry of research and technology would be dismantled and that the principal research council, the Centre National de la Recherche Scientifique (CNRS), would be weakened, has diminished as the fortunes of some of Chirac's more extreme science and education counsellors has waned.

Devaquet seems to be regarded as a "moderate" by those old-guard science policy makers who have had the chance to meet him, and his senior minister (Devaquet is "délégué" to the minister of education, René Monory) has even embraced some of the policies of his reforming socialist predecessor, Jean-Pierre Chevènement. Moreover, Monory's second-in-command, the more zealous Michèle Alliot-Marie, appears for the time being to have been put firmly in a more junior role. Another Chirac zealot, the medical scientist Gerard Milhaud, has so far been given no important post.

Devaquet's own powers and responsibilities were finally clarified last Friday. The ministry of research and technology will survive (in a somewhat truncated form) as the ministry of research and higher education, and there will be no "third tier" in the form of a prime ministerial office for research allocating budgets, as Milhaud and Alliot-Marie had been recommending. Thus for science, the buck will stop with Devaquet; but his ministry has lost control of a good deal of technology policy which has returned to the ministry of industry.

Devaquet's ministry nevertheless retains full budgetary control of the medical and agricultural research councils INSERM and INRA, against the initial advice of Chirac's radicals, and of ANVAR, the agency whose job is to find commercial applications for French academic research. Devaquet will, as expect-

ed, have full control of CNRS and of university research, which are no longer to be separate. Unlike his predecessor, however, the new research minister has lost control and finance for the whole industrial side of the atomic energy agency (CEA), the space agency (CNES), the agency for alternative energy and conservation (AFME), the telecommunications research agency CNET and the new £400 million science exploratorium, La Villette (except for fundamental sciences).

It is clear that policies on science and technology are being assiduously separated, which is expected of a government that believes in non-intervention in industrial affairs but which runs against the

recently published advice of the high scientific advisory council to the ministry of research, the Conseil Supérieure de la Recherche et de la Technologie.

Another danger for science may be represented by the threatened budget cut, which is one of the largest parts of a FF10,000 million (£1,000 million) saving in overall government spending promised by the new Prime Minister. Research alone is clearly more vulnerable than research in a package including industrial development, and the previous government's 1986 budget for research does stand high above the parapet, with a 4-5 per cent real growth previously planned against zeros in most other ministries. The likely cuts would wipe out this increase, while last week's eight per cent devaluation of the franc against the Deutschmark may take the real cut even deeper.

Robert Walgate

Academic copyright

Teacher/student conflicts

New York

THE American Chemical Society (ACS) is making a serious effort to regulate the disputes that may arise between academic researchers and institutions and the students who work in them on doctoral and other research programmes. The society is preparing a report on academic authorship arising from collaboration between faculty and graduate students. At the society's spring meeting last week, the ACS copyright committee discussed some of the ways in which disputes over copyright might be avoided.

This potentially contentious issue has hitherto attracted little attention from most scientific societies. On most college campuses, the copyright question is ethical rather than legal. Copyright law clearly considers any written work the sole property of the author, which reflects the liberal arts tradition that a research paper is an individual endeavour. But some researchers in the sciences contend that regardless of their involvement in the writing of a research report, their contributions in terms of experience, expertise and especially equipment merit co-authorship.

One ACS member wryly encapsulated his response to the graduate student who would remain independent: "Go get your beakers, go get your test tubes, purchase your NMR, then come back to me with your thesis — in 20 years". A student's contribution to his professor's paper is even more difficult to assess.

Based on responses from the deans of several prominent US graduate schools, the ACS committee has concluded that authorship abuses are infrequent, but suggests that professors and graduate students should outline their expectations in written agreements before beginning a re-

search project. Such measures will not necessarily bring about standard criteria for authorship. Frank Perkins, dean of the Massachusetts Institute of Technology's graduate school, thinks "there are a lot of graduate students who are troubled by the fact that some professors give them more credit than others" for the same contributions. He also points out that doctoral candidates may be reluctant to register formal complaints against their supervisors, since the complaints could easily boomerang.

The changing face of academic authorship first caught the spotlight last fall in a letter from Stanford University president Donald Kennedy to council members of the American Association of Universities (AAU). Citing an increased incidence of disputes between faculty members and students over credit for collaborative work, Kennedy called for "systematic discussion" of "the allocation of responsibility and credit for scholarly work", and of the forces that are driving research "towards a level of complexity at which it becomes difficult to determine responsibility of authorship".

AAU says the discussion Kennedy had hoped to provoke has not ensued. Neither have the courts hammered out precedents for copyrighting academic work, primarily because lucrative benefits have not been at stake. But a new kind of academic dissertation may soon change all that. Computer software developed by degree candidates can bring rewards more tangible than academic distinction, and could become a point of copyright contention. As software authorship becomes a matter of business, court decisions defining law may well set standards for academic authorship in general.

Karen Wright

US meteorology

NOAA faces rough passage

Washington

THE first launch from Cape Canaveral since the Challenger disaster on 28 January is scheduled for 1 May, when a Delta rocket will place a National Oceanic and Atmospheric Administration (NOAA) weather satellite into geosynchronous Earth orbit. The new satellite will restore NOAA's ability to monitor weather data from the east and west coasts of the United States simultaneously. But even as the Geostationary Operational Environmental Satellite-7 (GOES-7) goes into service, NOAA will be facing an uncertain future as Congress and the Reagan administration once again lock horns over the agency's budget.

In the past two weeks, NOAA administrator Anthony Calio has made three trips

Omnistar to rescue?

Washington

FUNDS for a second polar orbiting weather satellite and a \$97 million subsidy for the Earth Observation Satellite Company (EOSAT) were eliminated from the proposed budget for the National Oceanic and Atmospheric Administration (NOAA). But a novel plan is now under consideration that would allow NOAA to restore both programmes by flying them on the same spacecraft. The key to the plan is an engineering change proposed by EOSAT and approved by NOAA and the Department of Commerce last month. Under the new plan, the next generation of Landsat satellites will be flown on the Omnistar platform developed by RCA Corporation. Omnistar consists of an orbital platform to which modules are attached and has a design life of 20 years, significantly longer than the 5-year design life of the originally proposed satellites.

Omnistar would be launched by the shuttle, and its modules serviced by shuttle astronauts. Depending on precisely how Omnistar is configured, it could carry more instruments than just those needed by Landsat, which is why NOAA thought of combining Landsat with the polar weather satellites NOAA K-L-M, money for which has already been allocated, although no contracts have been let.

But critics of the plan fear crucial climate instruments will be left off the hybrid satellite. Francis Brotherton, chairman of the Earth System Science Committee, says that two instruments, the High Resolution Infrared Sounder and a sensor for measuring atmospheric ozone, are unlikely to be included in the new configuration, "a serious blow to long-range science". The proposed plan is now being considered by the White House Office of Management and Budget (OMB). Joseph Palca

to Capitol Hill to testify before congressional committees looking into NOAA's budget proposal for the coming fiscal year. For the past five years, NOAA has been on a budgetary roller-coaster. Each year, the President's budget proposal has contained significant reductions in the NOAA appropriation, and each year Congress has restored a large portion of the cuts. For the next fiscal year, NOAA's proposed operating budget is \$836.5 million, down 26 per cent from 1986. Support for ocean and coastal programmes is cut in half, for fisheries programmes by 40 per cent and for satellite programmes by 20 per cent. Gone entirely are funds for the National Sea Grant College Program, the National Undersea Research Program and the Earth Observation Satellite Company (EOSAT).

EOSAT was established by Congress to commercialize the Landsat system, and is a partnership between RCA Corporation and the Hughes Aircraft Company. Last September, EOSAT received a \$250 million subsidy for the construction of two new Land Remote Sensing Satellites (Landsats), with the understanding that additional subsidies would be provided until the new Landsats were launched. Cutting the money would be a "death knell" for EOSAT, says company president C.P. Williams.

It likely that funds for some of these programmes will be restored. At a Senate appropriations subcommittee hearing last week, Senator Lowell Weicker (Republican, Connecticut) vowed that programmes mandated by Congress but cut by the administration "will be funded again". But the self-imposed budget constraints of the Gramm-Rudman deficit reduction act may make it harder than in the past for Congress to restore those funds.

A NOAA administrator, Calio is in a very difficult position. A political appointee, Calio is obliged to support the budget handed to him by the White House Office of Management and Budget (OMB). But he must also seek to meet the concerns of Congress, which is clearly not satisfied with NOAA's budget. And unlike the National Aeronautics and Space Administration (NASA), NOAA is not an independent agency; Calio's actions must be approved by Secretary of Commerce Malcolm Baldrige. Shortly after President Reagan took office, a plan was proposed to merge the Department of Commerce with the Office of the US Trade Representative, splitting NOAA off as an independent agency, but that did not happen.

Calio became NOAA administrator in 1985, after serving as deputy to his predecessor John Byrne. While Calio's scien-

tific credentials do not match Byrne's, his administrative skills are superior. Calio spent 18 years at NASA in a variety of administrative capacities, and learned how to cope with Washington's political climate. Byrne, although favourably regarded by the academic community, did not promote NOAA's cause effectively. Tension between Byrne, Calio and associate administrator James Winchester over authority made matters worse. Now Byrne and Winchester are gone, and Calio plans to replace the position of associate administrator with a chief scientist who will be third from the top in the NOAA hierarchy.

The creation of a chief scientist post is one of the ways in which Calio has sought to enhance the scientific reputation of his agency. He has requested the National Academy of Science boards of atmospheric science and ocean studies to review NOAA's activities and suggest agendas for the future. The need for some federal agency to take the lead in collecting world climate data will be underscored by two reports soon to be released.

One, *Earth Science Research in the Civil Space Program*, prepared by the White House Office of Science and Technology Policy (OSTP) and awaiting approval by the Senior Interagency Group on Space, calls for the establishment of long-term global databases of environmental parameters to allow study of the Earth's evolution and civilization's impact on the environment. The report suggests that NOAA be given primary responsibility in collecting and managing those data.

Similarly, a report by the Earth System Science Committee of the NASA advisory committee sees NOAA as a key agency in keeping track of the enormous flow of data from Earth observing systems, a flow that will increase dramatically in the next decade. The OSTP report estimates that in the next decade, 10^{13} bits of data will be generated daily from space-based systems alone.

But Francis Brotherton, chairman of the Earth System Science Committee, has "very strong doubts" about NOAA's ability to perform the role. Because long-term measurements are crucial for forming an integrated picture of worldwide environmental processes, Brotherton believes entrusting so crucial a role to an agency so often buffeted around by strong budgetary winds could be disastrous.

Calio is aware of these difficulties, but nonetheless believes that NOAA can and should take on added responsibilities. He has worked hard to find ways of "doing more with less". He points proudly to TOGA as an example of NOAA's ability to involve other countries in US programmes. But Calio faces an uphill struggle in convincing the current administration of NOAA's importance.

Joseph Palca

US physics

Changing pattern in north-east

Boston

ALTHOUGH a large percentage of the US government's 1987 research and development budget is destined for defence-related projects, not all Strategic Defense Initiative (SDI) research is being supported by "new money". Physics researchers whose projects have been going on for years, says an analyst at the American Institute of Physics, are suddenly finding their grants recategorized "SDI".

This technology bleed between defence needs and physics research, particularly in astronomy, makes it difficult to tell whether defence monies are reshaping the pattern of university research. Both Dr Frank Pipkin and Dr Jerome Friedman, heads respectively of the physics departments at the Massachusetts Institute of Technology (MIT) and Harvard University, insist that government funds for research on their campuses carry no strings. Both universities stipulate that no professor can accept a grant for research which is secret.

That is why most objective-focused defence research is instead being carried out by industry as well as government laboratories which have become prominent since the Second World War in such fields as isotope separation and lasers. As a result, instrumentation at Lawrence Livermore and IBM, for example, con-

trasts temptingly with university laboratory facilities.

Harvard's instrumentation, says Pipkin, is "decrepit": especially since the arrival of computer-controlled instruments, the cost of updating laboratories runs into millions of dollars. Researchers who have chosen an academic setting, he says, must pick their physics problems carefully, accepting as a boundary condition their inability to compete in certain areas.

Dr T.K. Williams, staff physicist at the Department of Energy (DOE), differentiates between teaching instrumentation and government-funded research equipment. Two interdepartmental laboratories at MIT illustrate the gap: the National Magnet Laboratory, funded by the National Science Foundation (NSF), and the linear accelerator laboratory, funded by DOE, are both equipped with modern instruments and draw users from industry as well as academic institutions.

Although research funds may be string-free, university physics programmes must still dance to the tune of government grant agencies. As in the national trends, most of Harvard's 120 and MIT's 300 graduate students in physics are supported by research assistantships. At MIT, the sources are DOE and NSF, as at Harvard, as well as the National Aeronautics and Space Administration because of MIT's large astrophysics programme. Harvard will not be able to make good cuts in this support, says Pipkin; although nobody can yet predict the full effects of the Gramm-Rudman deficit reduction act, less support means fewer graduate students.

Meanwhile, the use of the legislative process to set aside funds for specific university programmes, as at Northeastern and Boston Universities, has been a cause of controversy. Academics at Harvard and MIT are unanimous in calling this a "distortion of funding" and "a dangerous trend", but do not want to be personally identified because they acknowledge that these "set-asides" are a way in which lesser programmes may be given a hand up in the competition for research funds.

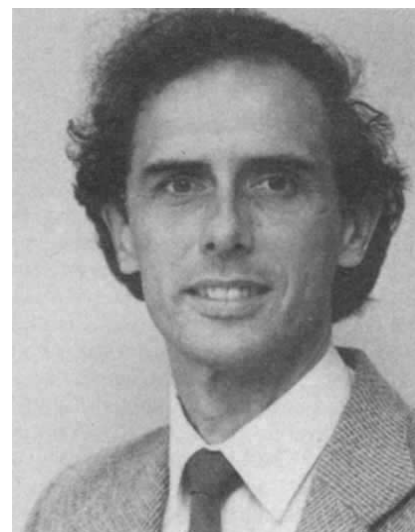
One of the beneficiaries is Boston University, whose president, Dr John Silber, is committed to transforming the graduate school programme. The university has, for example, committed itself to spend \$10 million over two years on its physics department, with continuing support of \$1.5 million a year. Undergraduate enrolment is expected to triple and graduate enrolment to double (from 55 to 100). Sixteen new faculty positions are scheduled to be filled over five years, but already four offers have been accepted and four are outstanding.

These increases presume that Boston

University can find a niche for itself by concentrating on fields such as computational physics and superstrings, which are not specialities at Harvard and MIT, and by fitting in between the "small size" of Harvard and MIT, plagued by "a faculty of old men with decreasing grants" (a charge Friedman refutes). Its new faculty members list the excitement of transformation as an important reason for coming to Boston University. **Elizabeth Collins**

Vancouver move

Dr John Schrader, the Australian immunologist, is to be the first director of the Biomedical Research Centre that is being established in Vancouver, Canada. Currently at the Walter and Eliza Hall Institute of Medical Research in Melbourne,



Schrader will take up his post in September and will have a staff of about 30 scientists. The centre has been set up by a company that is half owned by the Terry Fox Medical Research Foundation of British Columbia, a charity established by the province of British Columbia, and half by Wellcome Biotechnology Ltd, a subsidiary of Wellcome Foundation Ltd, the British pharmaceutical company.

The new centre has a research budget of £15 million over five years. Continued support will depend on its research record as well, no doubt, as on sales of Wellcome's interferon in British Columbia. The interferon is to be made in a cell culture facility, known as the Biomedical Processing Centre, that is to be built with Wellcome know-how but is owned by the Terry Fox Medical Research Foundation. The foundation, through its wholly owned company, Pacific Isotopes and Pharmaceuticals Ltd, has exclusive rights to make and market Wellcome's interferon in British Columbia. Eventually all parties concerned may benefit from the sales of other immunoregulators, a main focus of Schrader's present and future research. **Peter Newmark**

Benefits of inaction

Washington

THE US Congress won a bout last week in its tug-of-war with the administration over extramural research grants at the National Institutes of Health (NIH) — by doing nothing.

By failing to approve the administration's proposed \$77 million rescission to NIH's 1986 budget within a 45-day period, Congress increased the number of grants from 5,500 (the number sought by the administration) back to 6,100 (the number provided for by Congress).

The rescission, proposed on 6 February, would if implemented have cut \$53 million from the \$3,000 million total for extramural grants, as well as \$14 million from research into acquired immune deficiency syndrome (AIDS). NIH officials did not, however, seem unduly alarmed by the prospect: it is virtually unheard of for Congress to approve rescissions for a popular agency such as NIH. But because the number of grants has now jumped again, the cuts to individual grants now being negotiated (about 3 per cent below recommended amounts) will actually be greater than if the rescission had been approved.

Tim Beardsley

Netherlands research

OECD urges more investment

Waalre, The Netherlands

THE government of the Netherlands should increase its commitments to the quality and the strength of the research and development system. That is the most important of the recommendations by a group of examiners appointed by the Organization of Economic Cooperation and Development (OECD), whose report was published last week.

The document says that spending on civil research and development in the Netherlands has been a static proportion of Gross National Product (GNP) for several years, but that it has been a growing proportion of GNP in Japan, the United States and many other European countries. The OECD examiners say that the government of the Netherlands should seek to redress the balance if it wishes to retain its competitive position in the industrialized world.

The government contribution to research and development has been a declining proportion of the total in recent years, according to the report. Total expenditure on research and development by government and private companies is estimated to amount to 9,000 million guilders in 1986, or 2.15 per cent of this year's GNP. But while this represents a gradual increase over the 1.96 per cent of GNP spent in 1980, the government's share during this interval has increased from 0.93 per cent to only 0.96 per cent of GNP.

Although the spending of private companies has increased during this period, the OECD examiners point to the well-known circumstance that most company expenditure on research and development in the Netherlands is accounted for by only five private companies, typified by Shell and Philips, the electronics company. In 1986, it is estimated, these five multinational companies will account for two-thirds of the private company spending of 4,260 million guilders.

According to the OECD report, the Dutch government should be especially concerned to strengthen spending on fundamental research closely linked with new technology. It says that Dutch universities are lagging behind those in other comparable countries. Professor Eugen Seibold, president of the European Science Foundation and one of the examiners, points out that West Germany is spending relatively twice as much as the Netherlands on basic research.

Another examiner, Professor Christopher Freeman of the University of Sussex, urges that funds should not be distributed to universities by means of block grants but only after peer review.

The examiners also plead for more

spending on scientific equipment and on facilities such as information systems and libraries. The report suggests that there should be special efforts in the immediate future to make up for past neglect. Acknowledging that the system for the administration of science in the Netherlands is complicated, the examiners say that it is not easy to suggest changes, but agree with the decision made in 1983 that responsibility for technology policy should rest with the ministry of economic affairs.

According to Freeman, one of the advantages of this move is that it will help to educate economists more usually concerned with the question of macro-

economic growth in the relevance of technology to the question of growth and change. Although the minister of education and science in the Netherlands government will remain responsible for the coordination of technology policy, the involvement of the ministry of economic affairs will ensure that there is increased awareness of the relationship between technology and economic growth.

So much appears to have been recognized by the present cabinet, which has given the issue political priority. But because of the impending elections due on 21 May, and especially because the present coalition may not survive the polls, the ministers of economic affairs and of education and science have been reluctant to comment on the issues raised by OECD.

Casper Schuurings

Canadian research

Tax incentive scheme goes awry

Washington

THE Canadian Scientific Research Tax Credit (SRTC) programme, launched in 1983 to encourage investment in industrial research and development, has turned into a financial and political scandal. To a limited extent, the scheme functioned as intended, but it also provided an avenue for get-rich-quick schemes by financial entrepreneurs, and has cost Canadian taxpayers somewhere between \$900 and \$3,000 million in lost revenue.

Introduced as part of the 1983 federal budget, SRTC provided a 50 per cent tax credit for investment in research and development. The idea was to make cash available for new companies that would one day provide the backbone of Canada's high technology industry. Since many start-up businesses would not show sufficient profits to benefit from the credits, the government decided to make the credits transferable to other investors able to make use of the tax breaks. Under the arrangements introduced, small research and development companies would get much-needed infusions of cash from the sale of their entitlements while larger investors would save money on taxes. So everybody would be happy, and, in the long run, Canada would get a thriving new industry.

But problems surfaced rapidly. Lenders received tax credits for research in which they had no equity interest and so took little interest in how or whether it was carried out. Because there was no time limit for holding the SRTCs, the credits soon became negotiable financial instruments, and even items for speculation, with brokers making large profits from commissions on their sale.

In one common practice, called the quick flip, a large company or venture capitalist would lend a smaller research

and development company \$5 million, earning an entitlement to a \$2.5 million tax credit. The smaller company would then pay back half the loan, plus a brokerage commission of up to 12 per cent. The large company would get all its money back, split between cash and tax credits. The small company would end up with an infusion of \$5 million in cash, together with an obligation to spend \$5 million on research. Some companies never even attempted to do the research, but simply made off with the cash. Others tried what is called the double-flip, selling their research obligations to smaller companies willing to pay a higher commission for a quick infusion of capital.

The Canadian government called a halt to SRTCs in May 1985, after suspending parts of the programme in October 1984. Revenue Canada now has about 100 investigators performing audits on all tax returns claiming SRTCs, sending out teams of investigators when there is suspicion of fraud. Currently there are 58 companies under investigation. Revenue officials estimate that so far \$220 million has been recovered, but as much as \$900 million may have been lost forever. Analysts familiar with SRTCs reckon that the unrecoverable revenue may be much higher than the \$900 million the government acknowledges, perhaps as much as \$3,000 million. Supporters of the tax credit scheme nevertheless say the system could have been saved and abuses prevented with only slight modifications. By requiring SRTCs to be held for at least two years, schemes like the quick flip would have been impossible.

According to Guy Steed, associate director of research for the Canadian Science Council, the fallout from SRTCs has tarnished the national image of industrial research and development. **Joseph Palca**

Soviet space

Gagarin year tells more

SOVIET media comments on the twenty-fifth anniversary two weeks ago of Yuri Gagarin's first space flight by a human being have been memorable not only for the ritual honorifics of the occasion but for the interesting details they have provided of the early days of the Soviet space programme. No doubt this reflects the current policy of openness shown on these affairs.

Gagarin himself died in a jet crash in 1968, apparently during a refresher course on returning to the cosmonaut programme after several years of ceremonial duties. His place in the jubilee celebrations was taken by the German Titov, his back-up, and subsequently the second Soviet citizen in space. Addressing a cere-

monial meeting in Moscow's Hall of Columns Titov revealed that medical and biological research on the effect of rocket flight on living organisms was begun as early as 1951 — in other words, during the last months of the Stalin era. This is a remarkable statement, as, so far as is known, Stalin's only interest in rockets at that time was as missiles; moreover, Sergei Korolev, who under Khrushchev was to become chief designer of the space programme, was still, at that time, technically an unrehabilitated "traitor".

The cosmonaut training programme, Titov claims, began early in 1960, and after considering the work patterns of a number of groups, including divers, polar workers and mountaineers, it was decided to recruit air fighter-pilots, who, as Korolev put it, are "universal specialists". Two hundred pilots, Titov said, were initially selected, of whom just 20 got through the physical and medical examinations and six were short-listed for "intensive training" for the Vostok flight.

The same atmosphere of openness, at least on non-technical issues, pervaded the entire celebrations. In a press conference from on board the Mir space station, cosmonaut Vladimir Solov'ev admitted that some of the on-board Mir equipment was giving trouble, while another tele-

vision programme, starring cosmonauts Valerii Ryumin and Aleksei Lenonov, dealt with emergency situations in space during the past 25 years.

As far as future developments are concerned, however, the celebrations revealed little that had not already been announced by the Soviet planners or deduced by foreign Baikonur watchers. Mir station was praised as the "base port" for a future manned permanently operating scientific research complex, and there were numerous calls for increased international cooperation. The Apollo-Soyuz link-up of 1975 and the recent Vega-Giotto project were hailed not only as notable achievements in themselves but as possible precursors of the World Organization for Space Study and Research for Peaceful Purposes which the Soviet Union would like to see established. Addressing the ceremonial meeting, Academician Vladimir Kotel'nikov noted that the Soviet Union is now planning several "major international projects" including the training of Syrian and French teams to work aboard Mir.

Although, naturally, nobody would voice an objection on such an auspicious occasion, the fact that the French are to have a second turn must come as a disappointment to some of the Soviet Union's junior partners in Interkosmos, for whom there has so far been no suggestion of a second manned mission.

Vera Rich

Tamarins to go home

Tokyo

TEN golden-headed lion tamarins imported illegally into Japan three years ago are to be returned to Brazil. The announcement, made at an environmental committee meeting in the Diet, represents a considerable victory for the World Wildlife Fund branch in Japan, which has been quietly lobbying the government for the monkeys' return since last year.

The plight of the tamarins caught the media's attention when the director of the Rio de Janeiro Primate Center appeared on television during a private visit to Japan protesting that not only had tamarins been illegally imported into Japan but that they were not being properly looked after in the zoos that had subsequently bought them (see *Nature* 318, 200; 1985).

The golden-headed lion tamarin is one of the world's rarest primates, with perhaps as few as 200 remaining in their native habitat in southeastern Brazil. Those that came into Japan were sneaked in with false papers showing them to have been bred in captivity in Guyana, a falsehood that Japanese customs officials should easily have been able to spot. Captive breeding programmes are under way only at Rio de Janeiro and the zoo at Jersey in the United Kingdom.

Although the Japanese government has agreed to ensure that the tamarins will go back to Brazil, the World Wildlife Fund is likely to find itself with the bill and the responsibility for looking after the animals on their journey home. Negotiations are under way over compensation for the owners of the tamarins and over where they will go back to in Brazil. No problems are anticipated in shipping them: the Belgian government recently successfully returned without loss 16 animals illegally acquired by an animal dealer. Alun Anderson

US space

Titan blows up on pad

Washington

A TITAN 34D rocket carrying a military satellite exploded seconds after lifting off from Vandenberg Air Force Base in California last week, further crippling US space efforts already in disarray following the shuttle accident earlier this year. The explosion marked the second failure in a row for the normally reliable Titan rocket. A Titan 34D exploded several minutes into its flight last August, when a turbo-pump failed at the same time as a fuel leak developed.

The Department of Defense uses Titan rockets to launch military satellites too heavy for the smaller Delta and Atlas Centaur rockets. Titan flights have been put on hold until an Air Force accident investigation team headed by Colonel Nathaniel Lindsey completes his report on last week's incident. At this stage, the Air Force will not speculate on possible causes of the accident.

Although the Department of Defense mission was classified, there is much speculation that it was a KH-11 photo-reconnaissance satellite, the device believed to have been lost in the Titan accident last year. This leaves just one

functioning KH-11 in orbit, according to John Pike of the Federation of American Scientists, with no replacements available. Launched in December 1984, the remaining KH-11 is expected to last only until late 1987. The next generation of photo-reconnaissance satellites, dubbed KH-12, must be launched aboard the shuttle, raising fears that for a period of time the United States may be without space-based photo-reconnaissance ability. Although Pike believes the operational KH-11 will last until a new satellite can be launched, a gap in reconnaissance observations could have a negative affect on arms negotiations since US verification would be hampered.

The loss of the Titan rocket will place additional burdens on the already crowded shuttle schedule once flights resume. The National Aeronautics and Space Administration has already indicated that missions with national security objectives will have first claim to space aboard the shuttle. A new expandable launch vehicle being built for military payloads is not expected to be ready until 1988. Shuttle flights are expected to resume some time next year. Joseph Palca

Science in Nepal

Poor outlook for development

Kathmandu

NEPAL is trying desperately to emerge from centuries of backwardness by creating a base of science and technology to sustain its economic development. The King of Nepal himself, Birendra Bir Bikram Shah Dev, has lent his weight to the current enthusiasm for a technological leap forward. As chancellor of the Royal Nepal Academy for Science and Technology (RONAST), he has asked the academy to formulate a national policy on science and technology, including the creation of national laboratories, the co-ordination of research and the creation of a national information system. But in spite of the academy's links with sister academies in the region and with the newly-created Third World Science Academy in Trieste, not to mention the bilateral assistance of industrialized countries elsewhere, the obstacles are formidable.

The new laboratories to be set up in Nepal include a national instruments laboratory with assistance from Japan, a laboratory for research in hydrology and hydraulics with West German assistance and a planetarium at Kathmandu with help from India. The last of these projects fits in with the priority programme.

Using a \$30,000 grant from the International Development Research Centre of Canada, RONAST has also launched a project for increasing the awareness of the country's 16 million population of science and technology. RONAST has trained some 40 science journalists in the print and broadcast media and has launched a science features service. Its weekly radio programme on science topics has become instantly popular.

Nepal's annual science budget is a meagre \$300,000. In the absence of an industrial culture, there is very little applied research activity; the economy depends mainly on agriculture and tourism. The government's stress is on the development of appropriate technology for people who mostly live on mountains. The Rural Energy Centre for Appropriate Science and Technology (RECAST) has trained potters in the manufacture of smokeless stoves, 12,000 of which have now been installed in rural homes. The target by 1990 is 160,000 stoves.

RECAST is also converting the 25,000 traditional watermills by redesigning the turbine blades to double their efficiency. These watermills make use of the flow in streams and waterfalls to produce mechanical or electrical power for oil extraction or for dehusking foodgrains. Apart from biogas plants and solar water heaters, RECAST has developed technologies for making cement from rice husk and using bamboo for reinforcing concrete. Now it is

engaged on developing a suitable device for taking drinking water from the low-land streams to people living in the mountains. At present, the hill people not only trek for fuelwood but spend hours in carrying water uphill.

The biggest obstacle in building a scientific base in Nepal is its totally inadequate scientific and technical manpower. Up-to-date figures are not available, but according to 1982 statistics collected by the National Council of Science and Technology, the country has fewer than 4,200 scientists. The breakdown is agriculture 754, engineering 1,371, forestry 175, medicine 624, natural science 1,027, technology 104 and veterinary science 85. The total number of PhDs in Nepal is fewer than 100. The king personally presents a gold medal to anyone awarded a PhD.



A Nepalese water-mill, grinding maize.

Nepalese students wanting to go for higher studies have to go to universities abroad since the Tribuvan University, the country's only university, offers science courses only up to graduate level. The first batch of civil engineers from Tribuvan came out last year and the first batch of medical doctors will graduate next year. There is virtually no research in the university science departments and the laboratories remain shut most of the time. Said one senior professor, "I teach two hours a week and, during the remaining 40 hours, have no work". Very few students want to take up research because there are no jobs.

A senior official of RONAST admitted that lack of trained manpower is a major hurdle in the country's ambitious plan to build a viable infrastructure for research and development. It has been planned to set up a science campus, affiliated to Tribuvan University, in each of Nepal's

five development zones. But more important, according to some scientists, will be the reorganization of the university itself, where teachers and equipment are idle.

Nepal is a typical study of how massive development aid and imported science can fail to help a poor country to develop. For three decades, numerous international agencies along with their own advisers, experts and prescriptions for development have been steering Nepal on the road to progress and industrialization. Even today, 70 per cent of all health, agriculture, forestry, education and other development projects are foreign-funded and directed by overseas advisers estimated to number 480. But despite the massive scientific and monetary inputs from the developed world, most Nepalese continue to be poor, illiterate and unhealthy. Agricultural productivity has in fact declined and industrialization remains a dream. Foreign aid and the role of experts has been the subject of several debates and seminars in Kathmandu. One seminar in 1983 organized by the independent Integrated Development Systems (IDS) concluded that experts and advisers have their own existence without really trying to improve the skills of the natives, and that foreign-aided development projects mainly benefit the government bureaucrats associated with them.

According to P.K. Adikari of the National Commission on Population and one of the participants in the seminar, officials like to work in foreign-aided projects because they are tempted by the imported cars, foreign travel and prospects of scholarships and studies abroad. "The problems we have are peculiar to Nepal", says IDS director Dr Devindra Raj Panday; "No amount of foreign money or advisers can solve them."

One of Nepal's major assets is water. "Fifteen per cent of the world's hydropower potential is in our country", says Dr Kamal Shreshta, former member-secretary of the Royal Nepal Academy of Science and Technology (RONAST). "But we have not tapped it." He says that if foreign aid were diverted to building hydropower stations, Nepal could become a major exporter of electricity. All of Nepal's motor cars and buses could be run on electricity instead of petrol and diesel fuel, now all of it imported.

In Shreshta's view, the solution for Nepal is a development strategy based on exploitation of local resources using relevant technologies that can be mastered by the Nepalis. "Although we have many foreign-aided projects on energy, nobody bothered to develop a heater based on local fuels for warming homes in winter", says a RONAST scientist. Middle-class homes and government offices of Kathmandu presently burn imported kerosene in a burner marketed by the Japanese at £125 apiece.

K.S. Jayaraman

Diagnosis of Huntington's chorea

SIR—There has been a good deal of publicity in the past few weeks over whether or not the predictive test for Huntington's chorea (HC) should be made clinically available. The argument so far seems to have been between James F. Gusella and his colleagues in the United States and the Medical Genetic Unit at Oxford (see *Nature* 320, 21; 1986).

Dr Gusella's reasons for withholding the test from clinical use appear to be threefold: first, there is still a slight risk of heterogeneity; second, the error rate is higher than at first supposed, and has gone up from 5 per cent to a possible 10 per cent or even 15 per cent in some cases; third, we have as yet no idea of the effects of this test on people, and careful researching of counselling methods needs to be done in controlled situations before it becomes generally available.

It is not within a lay organization's brief to comment on scientific aspects, except to say that we understand that enough evidence has now been gained to make heterogeneity unlikely, and it could, anyway, never be entirely ruled out. A test that was 100 per cent reliable would seem to be preferable, but many of those who are currently at a 50 per cent risk of carrying the HC gene would rather live with a 10 per cent uncertainty about their future than the situation they are in now. On an exclusion test in pregnancy, we are told that the error rate is likely to be considerably less. The third point regarding counselling seems to be the critical one. People at risk to Huntington's chorea are no different from the rest of the population, except for their anxiety about inheriting HC. Most of them are quite capable of making up their own minds as to whether or not the error rate is acceptable for them, and whether or not they really want to have this test either on themselves or on a fetus. Full and skilled counselling is obviously vital, to make sure both that the facts are understood and that people have considered the repercussions of such a test if it were either negative or positive. Many will need continuing support and help afterwards, as will their families. Members of the medical and allied professions must be concerned about consequences and mistakes, but by being over-protective they are denying some people the right to take responsibility for their own decisions and lives.

On behalf of this association I would therefore like to state that contact with a large number of people at risk has shown that there is a diversity of opinion within this group, but a significant number would want to take the predictive test even in its present form (on a recent survey, 65 per cent of those already attending a genetic clinic), and would therefore support the

release of the test for clinical use providing that any unit administering the test was able to offer extensive pre- and post-test counselling both in the clinic and in the community.

SHIRLEY DALBY

*Association to combat
Huntington's Chorea,
108 Battersea High Street,
London SW11 3HP, UK*

SIR—As a lay person intimately involved with Huntington's chorea (affected husband, daughter, and three other children at risk) I can only applaud James F. Gusella and those at Massachusetts General Hospital who are so wise and sensitive to the many problems associated with the administration of the predictive test for Huntington's chorea. The test has never been given. No one knows how the results will affect individuals and their families. Until the pilot programme is initiated, concluded and evaluated, the moral and ethical issues, for the greater good, will not be known.

It has taken more than 100 years to come this far. To release the probe for clinical use before the initial pilot programme findings have been reviewed may be an error. Another year to prove that the test is in the best interest of all concerned does not seem unreasonable.

Again, I thank Dr Gusella and his colleagues for looking beyond the slides and test tubes to the individuals this test will ultimately effect.

NATALIE FRANK

*10 Louder's Lane,
Jamaica Plain, Massachusetts 02130, USA*

In defence of curiosity

SIR—In his comment¹ on the exchange of letters between Gusella and Watt *et al.*², John Maddox, while not denying the principle that discoveries are common property, puts unjustified restrictions on the application of this principle by presupposing a researcher's right "to demand a full account of the purposes for which the material is required" and by restricting acceptable purposes: "Mere curiosity is an insufficient basis for a request".

Two important points need to be made: first, discoveries, especially in the biological sciences, are frequently coincident with materials. Thus for researchers without access to the materials, the discoveries are sterile and worthless. Secondly, not only has curiosity been, at least historically, the impetus for the scientific process, but it also continues to provide, more than any granting agency, the wherewithal for its continued advance. Any retreat from the principle of free access to knowledge seems to be a retreat from the unrestrained

pursuit of knowledge. Scientific research is an expression of unrestricted *curiositas*.

The obligation of researchers to provide published materials must be absolute, tempered only by their overriding responsibilities to society. But without sufficient sanctions, some researchers may be tempted to restrict access to their materials for reasons of self-interest. We applaud the stand taken by other journals in demanding adherence to the principle of free availability and deplore the explicit denial of this principle by *Nature*. Whether the responsibility of Gusella to the patients of Watt *et al.* outweighs the right of Watt *et al.* to the probes discovered by Gusella is arguable. But the denial of the existence of the right of Watt *et al.* to the material is unjustifiable.

Reflecting on the idea of science and the role of curiosity in its history we must agree that "mere curiosity is indeed a sufficient basis for a request".

MILES BRENNAN

UTE HOCHGESCHWENDER

*Max-Planck-Institut für Immunbiologie,
Postfach 1169, D-7800 Freiburg, FRG*

1. Maddox, J. *Nature* 320, 11 (1986).

2. Watt, D.C., Lindenbaum, R.H., Jonasson, J.A. & Edwards, J.H.; Gusella, J.F. *Nature* 320, 21–22 (1986).

Paranormal theories

SIR—David Marks' Commentary on the paranormal (*Nature* 320, 119; 1986) portrays a marriage between the methods of scientific discourse and political debate. More disinterested observers might still regard the paranormal controversy as a valid exchange of ideas and refrain from attempts to end the discussion by attributing an arbitrary quality of decisiveness to their own work.

Experiments dealing with complex phenomena require the control of multiple variables and it is not unusual for such work to be vulnerable to criticism. As the burden of proof lies in the positive demonstration of paranormal phenomena, it is unlikely that this question will be easily resolved and we must then decide how much "lack of evidence" suffices to prove the non-existence of the alleged phenomena.

In the final analysis, however, the study of such questions may emphasize our humanity more than our objectivity, despite our wish to be impartial. To quote the author: "Beliefs... have an active life of their own and fight tenaciously for their survival. They tell us what to read, what to think, who to trust..." While this statement is certainly true of paranormal investigators, it is also true of others, including Dr Marks, ourselves and perhaps the alleged St Thomas.

JAMES A. SCOTT

DENNIS YOUNG

*85 Woburn Street,
Lexington, Massachusetts 02173, USA*

Hans Bethe on solar neutrinos

Whatever can account for the much smaller flux of neutrinos from the Sun than the theorists predict? Now, at last, there may be an answer.

NOBODY forgets that the doctrine that the Sun's energy is the product of thermonuclear fusion within the core of the star was the product, more than half a century ago, of the ingenious mind of the young emigrant physicist Hans Bethe, then newly a refugee from Hitler's Germany.

Bethe's theory has been, of course, a huge success; not merely does it successfully account for astrophysical phenomena other than the production of energy by stars, but it accounts splendidly for the properties of the Sun, the most closely observed of all stars, except in one respect: the output of neutrinos from the Sun, as measured by the experiments due to R. Davis Jr in which the conversion of chlorine-37 into argon-37 under the influence of the solar flux of neutrinos is found to be less than a third of what Bethe's followers would predict.

It would be a delightful happening if the puzzle of the missing neutrinos from the Sun proves to be resolved by Bethe's most recent scientific paper, an elegant piece of argument in *Physical Review Letters* for 24 March (56, 1305; 1986). Not often can a person hope to provide care and maintenance for a theory after the lapse of so much time.

Bethe's starting point is twofold, the discrepancy between the measured and expected flux of neutrinos from the Sun and what he calls "a very important paper" by two Soviet physicists, S.P. Mikheyev and A. Yu. Smirnov, said to have been presented at a conference at Helsinki in June last year. Bethe says that he has come to the same conclusion by different arguments, which is generous of him considering that, on the way, he corrects the sign of a crucial term in an equation which is essential to the physics of his result.

Several concepts enter, of which the simplest is that the conversion of chlorine-37 into argon-37 (which is detectable atom by atom because argon-37 is both gaseous and radioactive) is a means of recognizing electron neutrinos, the partner particles of ordinary electrons. But the partners of muons are different, and called muon neutrinos. Indeed, the interaction of a neutrino with a nucleus of chlorine-37 is the inverse of the radioactive decay of argon-37, which explains why muon neutrinos are not detected in Davis's tank full of chlorinated hydrocarbon.

The second concept, less familiar, is that of neutrino mixing, predictable since

the time when people became familiar with Dirac's theory of the relativistic fermion, taken as a credible fact of life since the early 1960s. The idea is that because neutrinos (unlike electrons) have no electric charge, those actually observed in laboratory experiments will be superpositions (in the quantum sense) of whatever neutrino states there may be. (Since there is a third generation of charged fermions, called tauons, there is also a third generation of neutrinos that may participate in the mixing.)

Hitherto, neutrino mixing has been one way of accounting for the missing neutrinos from the Sun. People have supposed that the neutrinos reaching the Earth are superpositions of electron and muon neutrinos, have reckoned that only the electron component of the compound state will be detectable and have gone on to guess at what admixture of the muon-neutrino state will account for the discrepancy; the snag has been that this estimate conflicts with that obtained from measurements near nuclear reactors (where it should be possible directly to measure the way in which the ratio changes with distance from the reactor, "oscillating" in the process).

What Bethe now does is to calculate the behaviour of neutrinos in the deep interior of the Sun, the novel step in his calculation and that of his Soviet precursors. On the journey from the Sun to the Earth, where there is not much of any kind of matter, the properties of the neutrinos are not much affected by external events. But in the interior of the Sun, where the density is high, there is matter enough with which the neutrinos may interact. But muons are scarce, so that only the electron neutrino part of the mixed state of the neutrino will interact. It is as if, says Bethe, electron neutrinos experience a potential energy dependent on the density (or the electron density) of the matter in their locality. But this feeds back to affect the mixing ratio, with the result that observable neutrinos are more like muon neutrinos in the interior of the Sun (where the density of matter is greater and so the potential energy of the electron neutrinos that much greater).

Another way of putting this is that the effective mass of the electron neutrinos (for energy is mass) is greater in the interior of the Sun than in empty space. But do neutrinos have mass? Maybe a little, perhaps just enough to close the Universe. And if they do have mass, the masses of

the electron and the muon neutrinos will be different while the masses of the measurable mixed states will be individually constant wherever they are measured, which leads quickly to the conclusion that in a region where the density of matter is great enough, electron neutrinos will be heavier than muon neutrinos if the opposite is true elsewhere. And that implies that within some critical radius within the Sun, electron neutrinos will be converted into muon neutrinos because they are more energetic, even though in empty space the energy balance goes the other way. In other words, there is a possibility that the electron neutrinos produced in the Sun's interior (predominantly by the conversion of boron-8 into beryllium-8) will be converted into muon neutrinos so effectively that they will hardly have a chance to reach the Earth.

With this conclusion, Bethe nicely turns the problem on its head, supposing that the conversion of electron into muon neutrinos in the interior of the Sun accounts for Davis's discrepancy, and going on to ask what that implies for the properties of neutrinos. The result is startling, if only because it shows how much an ingenious person can extract from a working hypothesis as simple even as this. "Now I propose to take this theory seriously" is how Bethe puts it.

One conclusion (which hangs on the assumption that neutrinos have some mass, however small) is that electron neutrinos whose energy is small enough will emerge as such from the interior of the Sun; putting in numbers, only one in eight of the neutrinos from the conversion of boron-8 to beryllium-8 will reach the Earth. This is more than enough to explain Davis's discrepancy, but also suggests that the gallium experiment planned by J.H. Bahcall should be less affected.

But the conclusion that stands out is Bethe's estimate of the neutrino mass. If the tauon neutrino is ignored, and the less massive of the observable neutrinos assigned zero mass, the heavier observable neutrino (in which the muon neutrino predominates) has a mass of only 0.008 eV. This is much less than the mass required to close the Universe. It is also less than that measured in the few experiments reporting a positive result. Bethe may well be right in claiming that astrophysics is a better way to a determination of the neutrino mass than any other kind of experiment.

John Maddox

Theoretical physics

Superstrings and supersymmetry

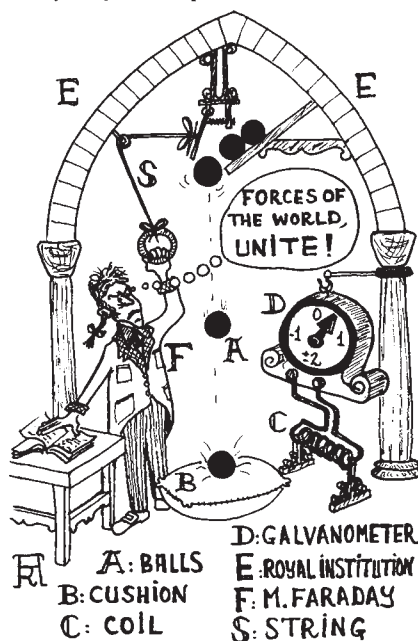
from Alvaro De Rújula

ONLY two years ago almost all elementary particle physicists considered the subject of 'superstrings' abstruse and irrelevant, perhaps because the number of space-time dimensions in which string theories can be consistently defined is somewhat unrealistic — either 10 or 26 or an astonishing 506. But in September 1984 Michael Green and John Schwarz published a paper¹ on the cure of certain diseases (called anomalies) in these elaborate theories. Overnight, most particle theorists shelved whatever they were thinking about (mainly 'supersymmetry' and 'supergravity') and turned their attention to superstrings. So far, none of these fashionable subjects has proved to have any convincing relationship with physical reality, and yet they have the irresistible power of addiction. Such a gregarious fascination for theories based almost exclusively on faith has never before charmed natural philosophers, by definition. What is the rationale for this seemingly unscientific revolution?

The key to understanding this fashion among particle theorists lies in the colossal potential attributed to superstring theory, defended by many as the 'theory of everything'. Many think that all the known and unknown elementary particles and forces, as well as the correct predictions for all their properties, will follow from superstring theory. Perhaps a more realistic claim is that superstrings may provide the first example of a consistent quantum theory of gravity, and of a unification of this force with all of the others: Michael Faraday's dream irresistibly come true.

Strings are the simplest generalization of conventional elementary point-particles: they are extended objects along a single space-like direction^{2,4}; they may vibrate and rotate; and their quantum 'normal modes' describe particles with different masses and spins. Some of these particles mysteriously turn out to behave as the graviton^{5,6} (the spin=2 carrier of the gravitational force) or as gauge bosons^{7,8} (the spin=1 mediators of electromagnetic, weak and strong forces). The 'old string' theory of particles of integer spin is known to be consistent only in 25 space dimensions and 1 time dimension, and has other even more appalling problems. But strings can be made supersymmetrical, that is, forced to describe particles with integer and half-integer spin^{9,10} related to each other in a specific and highly non-trivial fashion. Supersymmetrical strings, or superstrings, are consistent in a mere 10 space-time dimensions: a modest but significant step towards the real world.

Theorists are much less concerned about the four dimensions in which they live than about the internal consistency of their theories. Superstrings used to be disreputable because they contain anomalies — breakdowns by quantum effects of a conservation law obeyed by the theory at the 'classical' level — which are thought to be fatal. Some superstring anomalies are associated with gravitation, others with the remaining 'gauge' forces that the theory may encompass.



Michael Faraday pioneered the efforts to unify gravity and electricity by unsuccessfully attempting to observe electrical currents induced by accelerating bodies. Note his premonitory use of a string.

The explosive discovery of Green and Schwarz¹ was the cancellation of all anomalies for two particular cases of gauge groups: $SO(32)$ and $E_8 \times E_8$ (ref. 11). These groups are 496-dimensional, and describe that number of independent gauge bosons. But only 12 gauge bosons are required to describe the known electromagnetic, weak and strong forces: 1 photon; 3 weak intermediate vector bosons; and 8 gluons. We are thus presented with a beautifully unified theory, with 6 more dimensions than the real world and 484 extra carriers of interactions (see ref. 12 for review). The incantation invoked to avoid this embarrassment of riches is: "compactify!"

In general relativity, the local structure of space-time is determined by the interactions of the gravitational field with matter and with itself. Something a

thousandfold more elaborate is hoped to happen in a 10-dimensional superstring world. Six dimensions are expected to compactify spontaneously, that is, to warp around themselves into a finite space of unobservably tiny dimensions, leaving behind our flat four-dimensional reality. The geometry of the 'lost' six-dimensional space must be gruesomely contrived, if one is to explain why the relative strengths of the known forces are so different¹³, and why many fewer than 496 gauge bosons are observed¹⁴. Enormous progress in the understanding of compactification is needed for superstrings to become realistic and to move beyond the stage of a beautifully inspired game.

If strings are to offer a unified description of all interactions, their natural energy scale is that at which gravitation and the other forces between elementary particles are of comparable strength. This energy domain at which one could directly test string theories is some 16 orders of magnitude greater than the maximum energy of the largest existing accelerators. Only the very low energy limit of string theories can be tested in the foreseeable future. This requires no new experiments, say the superstring extremists, only further development of theory. If they are right, much of the fun of the discipline — to make empirically testable predictions and to bet on the results of experiments — would be forever gone.

It is presumably unreasonable to believe that we have already stumbled on the theory of everything. This does not mean that strings (whose beautiful properties are undeniable) may not survive as part of the 'truth'. But it may be wiser to admire the surprising potential of the natural world than to overestimate recent progress in theoretical physics. Michael Faraday was clearly aware of this in the 1850s, when he described his hopes to unify electricity and gravitation: "If the hope should prove well-founded, how great and mighty and sublime in its hitherto unchangeable character is the force I am trying to deal with, and how large may be the new domain of knowledge that may be open to the mind of man?" □

1. Green, M.B. & Schwarz, J.H. *Phys. Lett.* **149B**, 117 (1984).
2. Nambu, Y. *Copenhagen Symp.*, unpublished (1970).
3. Hara, O. *Progr. Theor. Phys.* **46**, 1549 (1971).
4. Goto, T. *Progr. Theor. Phys.* **46**, 1560 (1971).
5. Scherk, J. & Schwarz, J.H. *Nucl. Phys.* **B81**, 118 (1974).
6. Yoneda, T. *Nuov. Cim. Lett.* **8**, 951 (1973).
7. Neveu, A. & Scherk, J. *Nucl. Phys.* **B36**, 155 (1972).
8. Gervais, J.-L. & Neveu, A. *Nucl. Phys.* **B46**, 381 (1972).
9. Ramond, P. *Phys. Rev.* **D3**, 2415 (1971).
10. Neveu, A. & Schwarz, J.H. *Nucl. Phys.* **B31**, 86 (1971).
11. Gross, D.J., Harvey, J.A., Martinec, E. & Rohm, R. *Nucl. Phys.* **B256**, 253 (1985).
12. Green, M.B. *Nature* **314**, 409 (1985).
13. Candelas, P., Horowitz, G.T., Strominger, A. & Witten, E. *Nucl. Phys.* **B258**, 46 (1985).
14. Witten, E. *Nucl. Phys.* **B258**, 75 (1985).

Alvaro De Rújula is in the Theory Division at the European Laboratory for Particle Physics, CERN, 1211 Geneva 23, Switzerland.

Theoretical physics

Quantum tunnelling towards an exploding Universe?

from Malcolm J. Perry

THE discovery that superstring theory might enable one to form a consistent quantum theory of gravitation unified with the other fundamental forces — the strong, weak and electromagnetic interactions of matter — has generated an excitement in the theoretical physics community unparalleled since the invention of quantum electrodynamics (see Green, M.B. *Nature* **314**, 409; 1985 for a review of recent superstring theories). Nevertheless, string theory does not yet yield a complete picture of the physical world as we know it. One problem is that it predicts that the Universe has either 10 or 26 space-time dimensions, in contrast with the 4 space-time dimensions in which we seem to live. What is the fate of the remaining 6 or 22 dimensions?

The idea that space-time has more than four dimensions was originally explored by Th. Kaluza and O. Klein during the 1920s. These authors produced a unified classical theory of gravitation and electromagnetism in which space-time was five-dimensional, being the product of a circle of very small spatial extent and an ordinary four-dimensional space-time. In this theory, the fundamental unit of electrical charge e is related to the circumference l of the Kaluza-Klein circle by

$$l = \frac{1}{e} \sqrt{64\pi G}$$

where G is Newton's constant. Thus, in

this theory, $l = 2.3 \times 10^{-31}$ cm. The theory has the advantage of predicting that electric charge is quantized. To probe such short distance scales, and hence to observe them directly, would require incredibly large energies of about 10^{19} GeV. Thus, we need never worry about the fifth dimension in ordinary experiments. But this simple model turned out to be unrealistic because it predicts that charged particles must be very massive, again $\sim 10^{19}$ GeV. But a beautiful feature of the theory was that it was based purely on five-dimensional general relativity, only adding the assumption that space-time was partially wrapped up.

Today, a Kaluza-Klein theory refers to any model in which space-time is partially compactified, making some dimensions of small spatial extent. These small 'compact' dimensions will always be hard to observe directly, but they have profound indirect consequences. This compactification is presumably required in string theory. In the original Kaluza-Klein example the circle gave rise to electromagnetism. In more complicated examples based on $M^{(4)} \times \Sigma^{(d)}$, where $M^{(4)}$ is the large four-dimensional space-time and $\Sigma^{(d)}$ is a d -dimensional compact space, the isometry group of Σ manifests itself as the low-energy gauge group of a Yang-Mills theory (a relativistic point-field theory). Thus, in the original example, the iso-

metry group of the circle S^1 is $U(1)$. We can generate many gauge groups in this way, for example $SO(n)$ emerges if Σ is S^{n-1} , and $SU(n)$ emerges if Σ is CP^{n-1} . In all cases we get a Yang-Mills-type theory with a coupling constant g inversely proportional to the linear dimensions of the compact space. The difficulty with such theories is that there is usually no natural mechanism to make space-time look like $M^{(4)} \times \Sigma^{(d)}$, as opposed to $M^{(6)} \times \Sigma^{(d-2)}$ or $M^{(4+d)}$. If there are solutions that look four-dimensional on large scales, one can argue that such a space-time is appropriate and perfectly satisfactory. Such an anthropic argument is unconvincing because there are other equally plausible solutions that do not correspond to the Universe as we observe it.

So far, I have based the discussion on classical mechanics. Recently, E. Kolb and J. Frieman (*Phys. Rev. Lett.* **5**, 1435; 1985) have shown that when quantum mechanics is taken into account, such an argument loses all its appeal. They observe that although a space-time of the form $M^{(4)} \times \Sigma^{(d)}$ may well exist, a solution of the field equations looking like $M^{(4+d)}$ will also usually exist. $M^{(4+d)}$ is typically a $(4+d)$ -dimensional version of de Sitter space, a maximally symmetrical space-time, whose spatial extent is exponentially expanding on a timescale τ . Classically, a transition from one type of solution to the other is forbidden by the existence of a large potential barrier. In quantum-mechanical terms, parts of the Universe can tunnel from the desired state to a $(4+d)$ -dimensional de Sitter space. The probability of a transition per unit volume per unit time is calculated to be

$$\Gamma \sim \frac{1}{l_p^3} \exp(-c\tau^2/l_p^2)$$

where l_p is the Planck length (typical of gravitational phenomena) 1.6×10^{-33} cm and c is a constant. Both are determined by the detailed dynamics of the theory, but typically $1 \leq c \leq 100$ and τ is around 1 to 10 Planck times. (One Planck time is the time taken for light to travel one Planck length, and is 5.3×10^{-44} s.)

Thus, if c and τ have appropriate values, we can live in state which is $M^{(4)} \times \Sigma^{(d)}$ and can be reasonably certain that we will remain there for several times the observed age of the Universe. But it is possible that c and τ are such that Γ becomes so large that it is likely to tunnel into a state in which the Universe is $M^{(4+d)}$. Here, space would be $3+d$ -dimensional, expanding exponentially on the very short timescale τ . This unfortunate catastrophe would, like all quantum phenomena, occur completely at random. A calculation of c and τ in string theory then presents an interesting challenge. □

Malcolm J. Perry is in the Department of Physics, Joseph Henry Laboratories, PO Box 708, Princeton, New Jersey 08544, USA.

Strings, strings and, ultimately, strings

ONLOOKERS might be forgiven for confusion or even irritation at the apparently growing tendency of cosmologists and high energy physicists to invoke wrapping fibre when attacking fundamental problems. Here are the three types so far applied:

1. Strings — as in 'strings of galaxies'. Purely descriptive, although with implications for processes of galaxy formation (see Silk, J. *Nature News and Views* **320**, 12; 1986).
2. Strings — as in 'cosmic strings'. Loops or strands of 'vacuum defects' resulting from interactions between regions of different vacuum ground-state energy in the very early Universe as it cooled. These structures would be of sub-microscopic thickness and cosmic length (see Hogan, C. *Nature News and Views* **320**, 572; 1986).
3. Strings — as in 'superstrings'. Entities about 10^{-35} m long; the smallest

scale so far considered by theorists. Observable particles correspond to fundamental modes of superstring vibration. When set in 10-dimensional space-time certain types of superstring solve many problems afflicting previous attempts to unify gravity with the other forces. General relativity and the 'grand unification' of the strong and electroweak forces appear as 'low' energy approximations in this context (See de Rújula, A., opposite page and Perry, M.J., above).

Only the first type of string is known to exist but all are stimulating those in the game. For example, superstrings may change our ideas about transitions wherein forces decouple. An extremely speculative gleam in the eyes of true cat's cradlers is that strings (1) were precipitated with the help of strings (2) generated at a transition associated with strings (3). Philip Campbell

Visual system

Mapping of motion perception

from Oliver Braddick

How do we see the spatial form of a moving object? Although the visual system shows temporal integration over periods of the order of 50–100 milliseconds, moving targets do not appear smeared to anywhere near the degree that this implies¹. Yet the detectability of such objects as a function of exposure time suggests that information is integrated for at least this length of time². David Burr and colleagues, who performed this work, now propose³ that this integration can occur within spatio-temporal 'receptive fields' tuned to the target's particular direction and velocity of motion (see figure).

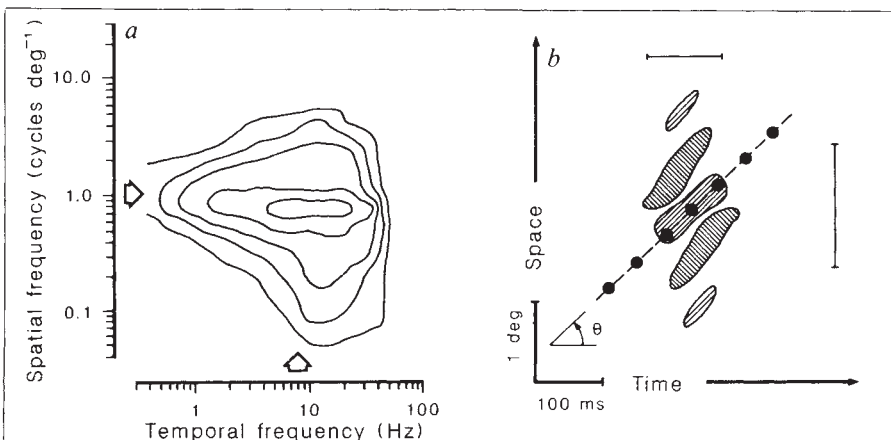
extend the masking method to the joint spatio-temporal domain, exploring how effectively each of a wide range of combinations of spatial and temporal frequencies can mask the appearance of a moving grating. Their results imply that moving gratings are detected by means of receptive fields, which can be described as parallel excitatory and inhibitory bands elongated obliquely in space-time, just as oriented bars and edges activate receptive fields whose form is elongated in a two-dimensional spatial plot⁴. Differently oriented fields in space-time then correspond to different optimal velocities, a

approximation to all of them.

Modellers of motion-sensitive detectors have been mostly concerned to explain how the visual system distinguishes different directions and speeds of motion. But the discussion by Burr *et al.*³ focuses on the problem of how we perceive the spatial form of a moving object. These authors propose that units with receptive fields tuned to direction and velocity of the target could integrate information so as to signal both the motion and the spatial properties of the target with precision. This elegant resolution of the problem does not entirely answer the question about the absence of perceived smear.

Presumably detectors tuned to other velocities, including zero, will summate signals over time in a way that does not keep step with the target's motion, but instead smears out its spatial representation. There would have to be some interaction between detectors, so that a sufficient signal from detectors tuned to the target velocity would suppress the signals indicating a smeared distribution of light from other detectors. Indeed, this problem is not special to motion, but is a general issue in understanding how perception is related to the information carried in parallel visual channels. The suppression of motion smear may be analogous to the way that the presence of sharp edges in a static image can suppress the processing of the 'smeared' low spatial-frequency components, as exemplified in the unrecognizability of block-quantized pictures — the 'Abraham Lincoln' effect¹².

The value of detectors that integrate information along the path of a moving object would presumably be to provide high-resolution information about its spatial form. Such integration could happen even if different spatial components of the image were never present exactly simultaneously. Burr¹³ and others^{14,15} have demonstrated fine vernier and stereo judgements in just such circumstances, and have argued that these 'interpolation' phenomena support the idea of inte-



a, Spatio-temporal sensitivity function of a motion detector in the human visual system. Contour lines represent increasing sensitivity, rising to a peak in the central region. The function was mapped psychophysically, using a test grating of 1 cycle deg^{-1} , moving at 8 deg s^{-1} (8 Hz) and a wide range of masking gratings. **b**, The same information in the domain of space and time. Receptive-field regions of hatching parallel to dashed trajectory line are excitatory and the others inhibitory. Dashed line, smooth motion; heavy dots, stroboscopic motion. Horizontal bar, length of time for which the field collects information; vertical bar, distance traversed during this time. $\tan \theta$, velocity. (Courtesy of D. Burr.)

Most neurones in the visual system are very responsive to moving images, and many respond selectively to a particular direction of motion and a particular range of speeds. Selective responses of visual neurones to spatial pattern are usually expressed in terms of their receptive-field profiles, or the spatial frequency sensitivity functions that can represent the same information in the Fourier domain. Because motion can be expressed as a pattern jointly in space and time, it is natural to try to describe the motion-selectivity of visual mechanisms in terms of a receptive field in space and time (or a sensitivity function which peaks for certain combinations of spatial and temporal frequency).

Psychophysical experiments using selective adaptation and masking are a well-established way of studying receptive-field structures in the human visual system, where they cannot be mapped directly⁴. In their new work, Burr, Ross and Morrone³

division among visual channels which appears to result primarily from differences in spatial response rather than temporal response (except for the special case of near-stationary stimuli).

A similar approach has appeared in two recent models of visual motion detectors^{6,7}. Both of these models contain, as a key element, linear summation of weighted signals from excitatory and inhibitory regions of a space-time receptive field. But linearity is not an essential postulate for the receptive-field concept; it just makes specification of receptive-field properties particularly straightforward. Model motion detectors that depend on non-linear interactions^{8–11} could still be considered in terms of spatio-temporal field maps. There are not yet very strong empirical grounds for preferring one of these formulations to another — indeed, there are strong suspicions that what is in the real nervous system might be some

1. Burr, D. *Nature*, **284**, 164 (1980).
2. Burr, D.C. *Proc. R. Soc. B*, **211**, 321 (1981).
3. Burr, D., Ross, J. & Morrone, M.C. *Proc. R. Soc. B*, **247**, 249 (1986).
4. Braddick, O., Campbell, F.W. & Atkinson, J. in *Handbook of Sensory Physiology, Vol. VIII: Perception* (eds Held, R., Leibowitz, H. & Teuber, H.L.) Ch. 1 (Springer, Heidelberg, 1978).
5. Hubel, D.H. & Wiesel, T.N. *J. Physiol., Lond.*, **160**, 106 (1962).
6. Watson, A.B. & Ahumada, A.J. *J. opt. Soc. Am.*, **A2**, 322, (1985).
7. Adelson, E.H. & Bergen, J.R. *J. opt. Soc. Am.*, **A2**, 284 (1985).
8. Barlow, H.B. & Levick, W.R. *J. Physiol., Lond.*, **178**, 477 (1965).
9. Reichardt, W. in *Sensory Communication* (ed. Rosenblith, W.A.), 303 (Wiley, New York, 1961).
10. Marr, D.C. & Ullman, S. *Proc. R. Soc. B*, **211**, 151 (1981).
11. Van Santen, J.P.H. & Sperling, G. *J. opt. Soc. Am.*, **A1**, 451 (1984).
12. Harmon, L.D. & Julesz, B. *Science*, **180**, 1194 (1973).
13. Burr, D.C. *Vision Res.*, **19**, 835 (1979).
14. Morgan, M.J. *Perception*, **5**, 187, (1976).
15. Fahle, M. & Poggio, T. *Proc. R. Soc. B*, **213**, 451 (1981).
16. Morgan, M.J. & Watt, R.J. *Vision Res.*, **23**, 997 (1983).
17. Welch, L. & McKee, S.P. *Vision Res.*, **25**, 1901 (1985).

Oliver Braddick is Lecturer in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

physiological effects of the enkephalins. The somewhat enhanced *in vitro* and *in vivo* potency of ketorphan compared with bestatin plus thiorphan also suggests an additional role for the dipeptidyl-aminopeptidase for which ketorphan has a very high affinity^{6,7}. In this regard, the use of tritiated inhibitors will be a useful way to locate the three enzymes in the central nervous system. Using autoradiography, enkephalinase is found in areas high in both enkephalin levels and mu- and delta-receptor binding sites, such as the striatum, periaqueductal grey matter, substantia nigra and spinal cord¹¹. What these inhibitors do not reveal is whether the increased availability of the enkephalins results from either a pool of tonically released peptide or peptide released as a result of a particular stimulus.

The potential of this approach to the production of physiological analgesia is high, particularly because effects mediated by the delta (enkephalin-preferring) opiate receptor may produce less of the problematical side-effects of the mu-preferring agonist morphine⁸. Furthermore, the results from animal studies discussed above indicate that the more inte-

grated measures of nociception may be altered by the inhibition of enkephalin degradation, whereas reflex tests seem, in general, to be insensitive. On the other hand, there is no reason to believe that the effects of the enkephalins are simple and enhancement of their synaptic availability and actions may result in tolerance, dependence, receptor changes and other compensatory mechanisms, leading to long-term changes in neuronal excitability¹². □

1. Roques, B.P. *et al.* in *Spinal Opioids and the Relief of Pain* 87 (eds Besson, J.M. & Lazorthes, Y.) (INSERM, Paris, 1985).
2. Giros, B. *et al.* in *Spinal Opioids and the Relief of Pain* (eds Besson, J.M. & Lazorthes, Y.) 57 (INSERM, Paris, 1985).
3. Dua, A., Pinsky, C. & LaBella, F. *Life Sci.* **37**, 985 (1985).
4. Kostelitz, H. *Proc. R. Soc. B* **225**, 27 (1985).
5. Waksman, G. *et al.* *Eur. J. Pharm.* **117**, 233 (1985).
6. Fournie-Zaluski, M.C. *et al.* *Eur. J. Pharmacol.* **102**, 525 (1984).
7. Fournie-Zaluski, M.C. *et al.* *J. med. Chem.* **28**, 1158 (1985).
8. Guillaud, G. *et al.* *Life Sci.* **33**, (1985).
9. Villaneuva, L. *et al.* *Brain Res.* **333**, 156 (1985).
10. Mendisohn, L., Johnson, B., Scott, W. & Frederickson, R.C. *J. pharmac. exp. Theor.* **234**, (1985).
11. Waksman, G. *et al.* *Proc. natn. Acad. Sci.* (in the press).
12. Woolf, C.J. & Wall, P.D. *Nature* **306**, 739 (1983).

A.H. Dickenson is in the Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK.

Ecology

The evolution of anti-predatory behaviour in zooplankton

from Felicity A. Huntingford and Neil B. Metcalfe

ALTHOUGH we often take it for granted that behaviour has evolved in response to natural selection, most of the evidence for this is indirect and circumstantial, involving deduction from the distribution of behavioural variants in modern forms. However, an article on page 746 of this issue¹ describes the rapid evolution of diurnal migration in zooplankton in response to stocking of lakes with fish and thus provides a notable exception.

One of the anti-predator responses of several vertebrate and invertebrate species²⁻⁴ is that of predator-induced shifts in feeding habitat, whereby animals accept a reduced rate of food intake to avoid predation. The diurnal vertical migrations which are such a conspicuous

feature of the behaviour of certain zooplankton species (ref. 5; see figure) could well reflect similar selection pressures⁶. Food resources are often more abundant in the upper water layers but, because of higher light levels, risk of predation by visually hunting planktivorous fish is also greater. During the day, the cost of increased predation outweighs the benefits of high rates of food intake and plankton migrate into deeper, safer but less productive water. This behaviour is most marked in species and age classes that are particularly vulnerable to predation^{5,6}. Some cases of predator-induced changes in feeding site involve short-term, reversible behavioural adaptations to variable local conditions in animals with relatively

advanced behavioural capacities. The vertical migration in zooplankton, however, is more likely to be a genetically pre-programmed behavioural change.

Gliwicz's data come from studies of the day- and night-time distributions of juveniles of an annual copepod species (*Cyclops abyssorum*) in a series of isolated lakes in the Tatra mountains of Central Europe. Some of these lakes contain long-standing populations of planktivorous fish, others are free of such fish and yet others have been artificially stocked, mainly within the past 25 years. Vertical migration only occurs in those lakes that contain planktivorous fish and this behaviour is more marked in lakes with long-established fish populations than in recently stocked lakes.

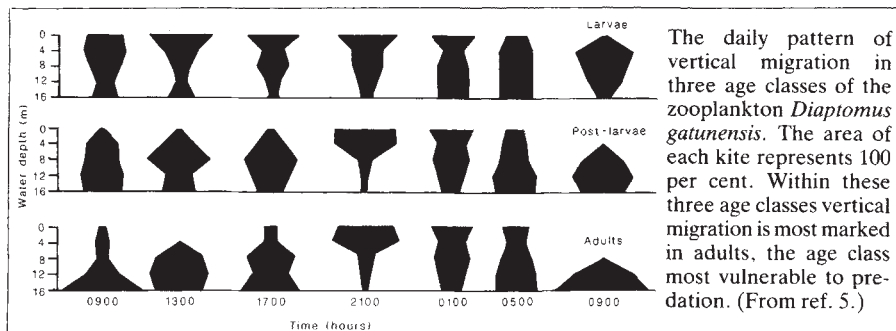
The lakes differ in other respects, such as visibility and depth, and so although compelling, these results still provide only circumstantial support for the so-called predation hypothesis. But Gliwicz's study provides another, uniquely valuable, kind of evidence, in the form of data on the migratory patterns of *C. abyssorum* in one particular shallow lake at several different stages in a programme of stocking with planktivorous fish.

Twelve years after stocking began in this lake there was little evidence of diurnal migration, but now, some 23 years later, the plankton show strong migration away from the water surface during the day, burrowing into the substrate. The fact that migration only appeared after the plankton had been exposed to predation for several generations suggests that this is not a short-term, phenotypic response to the presence of predators. The rapid appearance (in evolutionary terms) of the migratory response is ascribed to the fact that, even in lakes with no predators, some individuals show small diurnal movements, so that right from the start material was available for natural selection to work on.

Thus Gliwicz has tracked the gradual evolution of migratory behaviour in response to an artificially imposed increase in predation pressure. His study sheds light on the way natural selection moulds anti-predator responses, and so is comparable with the classic work on the occurrence of light and dark forms of the peppered moth (*Biston betularia*) which showed how the relative frequency of morphological variants change in response to a change in predation risk. □

1. Gliwicz, M.Z. *Nature* **320**, 746 (1986).
2. Sih, A. *Ecology* **63**, 786 (1982).
3. Stein, R.A. & Magnuson, J.J. *Ecology* **57**, 751 (1976).
4. Werner, E.E., Gilliam, J.F., Hall, D.J. & Mittelbach, G.G. *Ecology* **64**, 1540 (1983).
5. Zaret, T.M. & Suffern, J.S. *Limnol. Oceanogr.* **21**, 804 (1976).
6. Stein, R.A. in *Predator-Prey Systems in Fisheries Management* (eds Stroud, R.H. & Clepper, H.) 343 (Sport Fishing Institute, Washington DC).

Felicity A. Huntingford and Neil B. Metcalfe are in the Department of Zoology, University of Glasgow, Glasgow G12 8QQ, UK.

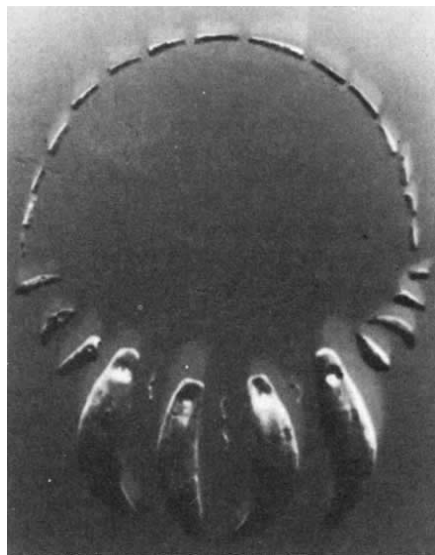


Archaeology

Rediscovering French ice-age art

from Randall White

To achieve new insights into the archaeological record, it has become almost as important to discover and reanalyse that which archaeologists have lost, ignored or misplaced as it is to recover new examples of materials lost or discarded by prehistoric peoples. In 1980, I began a systematic inventory of French Upper Palaeolithic (35,000–12,000 years ago) materials in American institutions. This has revealed a collection probably more significant than any other from this period outside France. It includes nearly 100 pieces of art; about 500 bone and antler implements; over 200 perforated beads, teeth and shells; and the partial skeletons of at least two late Pleistocene humans



Necklace as reconstructed by Collie.

In the 1920s, the Logan Museum of Anthropology at Beloit College, Wisconsin, launched an archaeological expedition to southwestern France led by George Collie and his assistant Alonzo Pond. The expedition made substantial excavations at two important Palaeolithic sites in the Vézère Valley: Abri Cellier¹ and Rocher de la Peine. The collections, still housed at the Logan Museum, contain several thousand stone tools and many worked bone and antler tools that provide information about manufacturing processes and the chronology of the sites. Although the stratigraphic analysis on which these excavations were based does not conform to modern standards, our meagre understanding of both these sites is much enhanced by these new-found collections. For example, there are about 250 worked bone and antler objects from the Aurignacian (33,000–28,000 years ago) and Gravettian (29,000–21,000 years ago) levels of the Abri Cellier including incised bird

bones that came to be known as hunting tallies, and two anthropomorphic pendants.

One of the few intact necklaces from the Magdalenian (18,000–12,000 years ago) was found at Rocher de la Peine (see figure). It is composed of three perforated bear canine teeth, one perforated and incised lion tooth, various other smaller teeth and many pierced shells from the Atlantic Coast, 150 km away. Also recovered were the complete mandible and fragmentary maxilla of an adolescent human. The original excavators thought it likely that the necklace was part of some burial furniture.

Collie and Pond purchased several large collections, the most spectacular being that of 40 engraved limestone blocks from the Magdalenian site of Limeuil, part of a larger collection of about 200 blocks recovered by Capitan and Bouyssonie in the 1920s (ref. 2). But several blocks from Capitan and Bouyssonie's collection were

never published because they were indecipherable or 'poorly done'.

Clearly, the present-day collection in France is incomplete and cannot be analysed in isolation from the Logan collection. Moreover, a cursory examination of the published blocks shows that Capitan and Bouyssonie's original analysis² left out most of the non-animal and heavily superimposed images. In the Logan collection there are several fragments of blocks which can be pieced back together with others at the Logan Museum and at the Musée des Antiquités Nationales in Paris. An international effort to fit together the isolated limbs and heads of animals to make larger images is now under way.

Pond also purchased material from the Abri Blanchard, the richest known site for the earliest Aurignacian art and ornamentation. Surprisingly, the Logan Museum holds the 'necklace' recovered and described by Didon³. This necklace is a series of nearly 150 perforated antler, bone and stone beads and 5 pendants, some of them incised and decorated. The most interesting pendant was described by Didon as an ivory fish (but is probably a seal or dolphin), which has an intricately patterned series of dots drilled on its back and sides,

Growth factor A-chain back in train

As soon as the limelight of oncogenes shone on the B-chain of platelet-derived growth factor (PDGF), the A-chain was cast in such a minor role that it was in danger of being forgotten. But after starring in two recent papers, one on page 695 of this issue and one published earlier this year (*Nature* 319, 511; 1986), the A-chain is back in prominence.

That highly purified PDGF contains the two chains is undisputed. And that the chains are present as dimers is equally certain. What has not been settled is whether PDGF comprises A–B dimers or a mixture of A–A and B–B dimers, and what the contribution of each chain is to the biological activity of PDGF.

Soon after the inferred polypeptide sequence of the *sis* oncogene of simian sarcoma virus and the actual polypeptide sequence of the B-chain of PDGF turned out to be the same, there followed evidence that the B-chain-like substance secreted by cells that have been transformed by the virus mimics most, if not all, of the properties of PDGF, including its ability to stimulate fibroblasts to divide. Therefore little role seemed left for the A-chain.

Matters changed early this year when Carl-Henrik Heldin *et al.* succeeded in characterizing from an osteosarcoma cell line a substance that had the activity of PDGF yet turned out to be the A–A rather than the B–B

dimer. The same group has now followed up with the DNA sequence of the precursor of the A-chain together with evidence of its expression in several tumour cell lines. These data, together with the clear demonstration that the A-chain gene is not only distinct from the B-chain gene but is on a different human chromosome, are presented in this issue (*Nature* 320, 695; 1986).

The data allow a detailed comparison of the structures of the A- and B-chains and their precursors. More interestingly, they show that the PDGF-like activity produced by most investigated tumour cell lines is probably A-chain rather than B-chain, although some cell lines transcribe both genes and perhaps secrete a mixture of the polypeptides.

This raises the question of whether the PDGF-like growth factors shown to be produced by an increasing number of normal cell types is A-chain rather than B-chain or a mixture. Indeed, as the dimers of either chain seem to have similar biological activity, why should there be two? Are there subtle differences in their activities? And is the activity of the A–B dimer different again — if, that is, it ever exists? These and related questions should soon have answers now that the A-chain and reagents for detecting it are in the hands of cell and molecular biologists.

Peter Newmark

and is comparable to the Aurignacian plaques from Blanchard and Lartet already studied by Marshack⁴ and thought by him to carry lunar notation. More importantly, it links the earliest art of south-west France with the similar and contemporary ivory statuettes from Vogelherd in Germany, and is further evidence for the complexity and sophistication of artistic expression in its earliest manifestations.

Armand Viré undertook massive excavations at four sites near Lacave, France, at the start of this century. These excavations, at the sites of Crozo Bastido⁵, Crozo Gentillo^{6,7}, the Grotte de Lacave and the Rivière de Tulle shelter⁸ yielded a series of spectacular assemblages dating to the Solutrean (22,000–18,000 years ago) and the Magdalenian (18,000–12,000 years ago). These assemblages, long considered to have been lost, were also purchased by Pond and are now at the Logan Museum. They include many examples of art and ornamentation, including pierced shells and a shark's tooth that could only have come from the Atlantic coast 200 km away. Apart from abundant Magdalenian

and Solutrean stone tools, there are hundreds of bone and antler artefacts, including several harpoons, dozens of bone needles, a decorated pierced baton and a huge number and range of carefully worked antler objects, some of them nearly two feet long.

The careful reanalysis now in progress will improve our present understanding of the Upper Palaeolithic in the Perigord and Haut-Quercy⁸, although there will always be limitations imposed by some rather poor stratigraphic work. □

1. Collie, G. *Bull. Logan Museum* 1, (Beloit, Wisconsin, 1928).
2. Capitan, L. & Bouyssonie, J. *Limeuil: son Gisement à Gravures sur Pierres de l'Âge du Renne* (Nourry, Paris, 1924).
3. Didon, L. *Congrès International d'Anthropologie et Archéologie Préhistorique* 6, 337 (1912).
4. Marshack, A. *The Roots of Civilisation*. (Weidenfeld & Nicolson, London, 1972).
5. Viré, A. & Teulière, C. *L'Anthropologie* 37, 449 (1927).
6. Viré, A. *L'Anthropologie* 16, 411 (1905).
7. Viré, A. & Niederlander, A. *Bulletin de la Société Préhistorique Française* 18, 269 (1921).
8. Lorchblanchet, M. in *La Préhistoire Française* (ed. De Lumley, H.) 1189 (CNRS, Paris, 1976).

Randall White is in the Department of Anthropology, New York University, New York 10003, USA.

Electron microscopy

Coulomb explosions in metals?

from A. Howie

IN comparison with bulk material, small particles containing less than a few hundred atoms have several anomalous properties, including low melting point, modified lattice parameters, multiple twinning and other structural defects or variations. These effects are of great interest in basic science as well as in crystal growth and catalysis. So far most quantitative studies of them have relied on mass spectrometry or diffraction observations, but direct imaging by high-resolution electron microscopy should in principle provide much more detailed information about individual particles. Recent use of direct imaging with real-time video recording reveal some intriguing and unexpected phenomena related to the high current density of the electron beams used.

The most striking results have been obtained on 2-nm diameter gold particles by Iijima and Ichihashi using a specially constructed, clean vacuum system at the Research Development Corporation of Japan. Preliminary images appeared last summer (*Nature News and Views* 315, 628; 1985) and a more detailed account has now been published (Iijima, S. & Ichihashi, T. *Phys. Rev. Lett.* 56, 616; 1986). Iijima and Ichihashi use a 120-keV electron beam at a current density of 200 A cm⁻² (1.3×10^5 electrons per Å² per second) to obtain a recording speed of 60 frames per second. The most noticeable feature of their results is that a given particle

can be seen to jump intermittently between several different states — single crystal, decahedron, icosahedron and complex multiple twin. Sometimes these states seem to be quasi-stable, persisting for a few tenths of a second and exhibiting, for example, local details of surface facets that minimize the energy. At other times, the particle is clearly highly excited and appears to be molten or spinning on the silicon dioxide-covered silicon support.

Although these abrupt changes of overall structure seem to be suppressed in larger (10-nm diameter) gold particles, equally startling and sudden convulsions of surface structure have been observed for particles of this size by J.O. Bovin *et al.* (*Nature* 317, 47; 1985) using a 300-keV electron microscope and a beam intensity of 20–30 A cm⁻². Their images show what appear to be atomic clouds of gold atoms outside the particle that interact with individual atomic columns viewed in projection at the particle surface and were immediately recognized as both interesting and puzzling (see *Nature News and Views* 317, 16; 1985).

The generation and dissipation mechanisms for electron-beam excitations in small particles are reasonably clear in principle, although not perhaps at the level of quantitative detail to answer unequivocally all the questions raised by these observations. The most probable process, the excitation of valence electron-

density oscillations (plasmons) has a characteristic energy-loss quantum of typically 20 eV and would involve one or two per cent of the fast electrons passing through or near the particle. These excitations can decay radiatively, or more probably by emission of a secondary electron, although the latter process can in some cases be inhibited if the particle is not sufficiently well-grounded through its support and becomes positively charged. Charging effects do indeed appear to be relevant because Iijima and Ichihashi find that the small particle motion becomes more sluggish when conducting supports of amorphous carbon are used. Plasmons, however, are generated at such a high rate (10^5 per second on each particle) that, on the timescale of video recording, one might expect an effective equilibrium with an associated temperature rise of roughly 100°C for the atoms.

The abrupt, kaleidoscopic image changes observed in these experiments are probably caused by some relatively much rarer event such as inner-shell excitation. Such collisions can involve not only the transfer of considerably larger amounts of energy and momentum to the particle but also substantial charging effects via the Auger cascade process. This process occurs when the energy released by filling of the inner-shell vacancy causes injection of two or more less strongly bound electrons from the atom. The resulting Coulomb explosion, a long-established concept in the theory of particle track formation, has recently been identified in various electron-beam desorption effects (Knotteck, M.L. *Nature* 291, 452; 1981 and *Rep. Prog. Phys.* 47, 1499; 1984). It may also play a key role in electron-beam ionization damage in aromatic organic crystals, where K shell ionization rather than valence excitation (Howie, A. *et al. Phil. Mag.* B52, 751; 1985) is the crucial event.

At the other extreme of scanning transmission electron microscopy, where beam-current densities in excess of 10^5 A cm⁻² can lead to dramatic nanometre-scale hole-drilling in normally stable inorganic materials (Salisbury, I.G. *et al. Appl. Phys. Lett.* 45, 1289; 1984), the Coulomb explosion has also been implicated. Hitherto it has been generally assumed that such effects would be very rapidly screened-out in metals, but the special features of small particles may make them more susceptible to this rather dramatic type of excitation. By looking for coincidences between the sudden changes of particle structure and characteristic energy losses associated with inner-shell excitation, it may be possible to examine this hypothesis more closely. □

A. Howie is Reader in Physics at the Department of Physics, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK.

Neurochemistry

Is GABA the neurotransmitter for some photoreceptors?

from Stephen Yazulla

A SERIES of papers published last year in *Nature* and discussed in these columns¹ caused major revision of a long-held hypothesis about photoreceptor transduction. Now, again in *Nature*, photoreceptors are the focus of another controversy. Nishimura *et al.* on page 753 of this issue² report that the synaptic terminals of about 25 per cent of the rod and cone photoreceptors in the monkey retina show immunoreactivity to antisera against γ -aminobutyric acid (GABA) and its biosynthetic enzyme L-glutamate decarboxylase (GAD). These authors suggest that GABA may be a neurotransmitter in a select population of rod and cone photoreceptors. If correct, the report is highly controversial: either it will require a major revision of our views on photoreceptor synaptic physiology; or, if incorrect, it means that the use of immunocytochemical labelling as a technique to localize neurotransmitter systems in neural tissue must be questioned.

Vertebrate photoreceptors constantly release transmitter in the dark. The effect of light is to hyperpolarize the photoreceptor, with a consequent reduction in transmitter release³. There is general agreement that the transmitter(s) of vertebrate photoreceptors are depolarizing (excitatory) to horizontal cells (second-order neurones) — a leading candidate for such a transmitter is the excitatory amino acid L-glutamate^{4,5}. GABA, however, tends to function as a hyperpolarizing (inhibitory) neurotransmitter in many parts of the nervous system, including the retina⁶. Thus, the implication of GABA in photoreceptor transmission by Nishimura *et al.*² appears diametrically opposed to the generally accepted hypothesis, which is founded on a very large database obtained by electrophysiological, biochemical and pharmacological means in the retinas of many vertebrate species⁷. It is unfortunate that the present data were obtained in monkey retina, a preparation for which there are no supporting physiological data, nor are there likely to be any in the foreseeable future.

To establish a chemical as a neurotransmitter, several criteria must be fulfilled, of which the demonstration that the candidate and its biosynthetic enzyme are present, as suggested in the present report, are only two. Primate retinas, because of cost, supply and viability *in vitro*, are much less amenable to the kinds of manipulations required to obtain supportive physiological data needed to satisfy crit-

eria such as measurement of transmitter release and physiological action on postsynaptic neurones. Herein lies the value of non-primate and non-mammalian retinas for the study of retinal function.

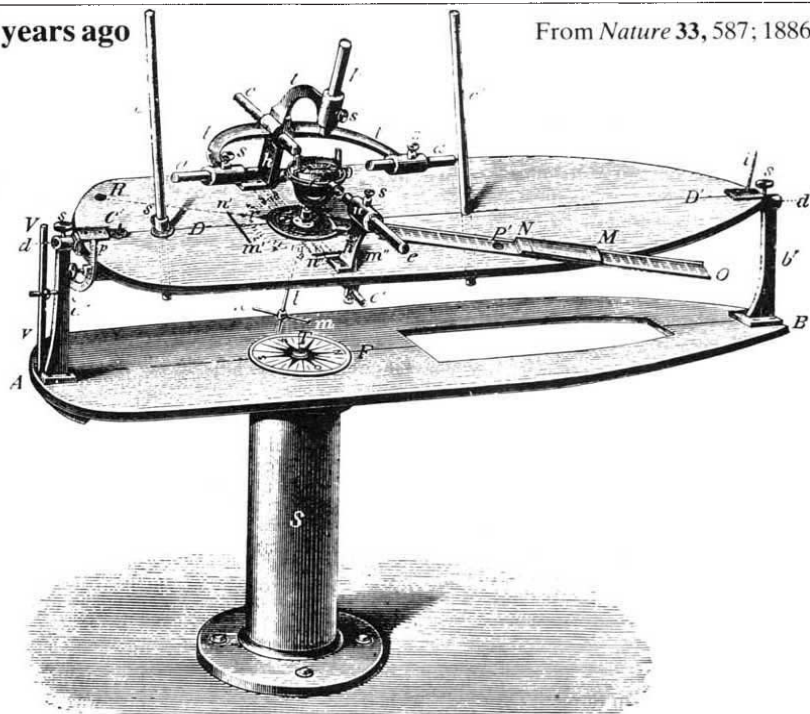
A further issue raised by the new report involves the validity of immunocytochemical protocols to mark or label neurotransmitter-specific pathways. There has recently been an explosion in the number of published papers that use immunocytochemical localization techniques to identify putative transmitter-specific neurones and pathways. Despite the care with which many antisera are produced and characterized, there are a host of possible artefacts that can lead to false-negative or false-positive results. Artefactual staining can be produced, for example, by cross-reactivity of the antisera with tissue proteins, impurities in the antisera and

nonspecific staining of necrotic or damaged tissue.

The data of Nishimura *et al.* are adequately controlled for such artefacts, with reliable and apparently valid staining in a class of neurones in which there is no other evidence for GABAergic transmission. If one chooses not to accept the data of Nishimura *et al.*, then doubt must be cast on all existing published immunocytochemical data for which adequate controls appear to have been done. The degree of such concern, which does exist, seems to depend on the transmitter system under study. For each system, many different antisera are available from different laboratories that use different protocols; they can be directed against the degradative or synthesizing enzymes of the transmitter, its receptor molecule or the transmitter itself. Thus, details in the localization of a transmitter system often differ depending on the antisera used, although the different antisera were presumably raised against the same antigen.

As an example, the GABAergic system in the retina has been studied using antisera against the GABA biosynthetic enzyme GAD. There are at least six GAD antisera currently in use which were

100 years ago

From *Nature* 33, 587; 1886.

THE deviations of the compass produced by the iron used in the construction of wooden ships was a source of considerable perplexity to the navigators of the last and early part of the present centuries; and no sooner were these difficulties overcome than the building of ships entirely of iron commenced. One of the most important contributions to magnetical science as regards iron ships was made by Sir George Airy (late Astronomer-Royal) in a paper published in the *Phil. Trans. Royal Society* for 1830 who invented the system of correction by magnets and soft iron, which is universally practised in the present day. This system of correction, coupled with the analysis described in the "Admiralty Compass Manual," provides the means of correcting a compass even in position on the 'tween decks of our armoured ships of war. Different forms of models have been used for instruction in the magnetism of iron ships considered. The illustration above is a woodcut of the original model taken from Paper No. 2 of the *Archiv der Deutschen Seewarte*, VI. Jahrgang, 1883, where an account of the experiments to be made with it is given in full detail.

raised in different animals against GAD extracted from the brains of different species (catfish, mouse, sheep, rat and cow). These GAD antisera differ in the degree of cross-reactivity with GAD from different species and probably recognize different isozymes of GAD. Although the overall labelling patterns in the retina are similar for each GAD antiserum, notable differences exist. An analogous situation is also found to exist for the many types of GABA antisera.

The GAD antiserum used by Nishimura *et al.* was raised in sheep against rat brain GAD²; it is one of the most widely used GAD antisera in the retina. However, this is the only report of photoreceptor labelling with this antiserum and there are no other reports of labelling with any of the GAD or GABA antisera in use elsewhere. Is it possible that 'GABA' labelling of photoreceptors is so alien to current thinking about photoreceptor physiology that any staining of photoreceptors may have simply been ignored? I suspect that those of us who perform GAD and GABA immunocytochemistry will take a closer look at old slides as a result of the new work by Nishimura *et al.*

Despite the orderly arrangement of photoreceptors at the distal margin of the retina, the photoreceptor transmitter(s) has not been identified. A leading candidate is L-glutamate, but it is so ubiquitous in nervous tissue that its involvement in photoreceptor transmission has been difficult to prove. GABA is synthesized from glutamate via GAD and both glutamate and GAD are components of the GABA shunt, a metabolic reaction that serves to stabilize GABA and glutamate concentrations. It is conceivable that the immunocytochemical localization of GABA and GAD in photoreceptor terminals is more closely related to glutamate/GABA metabolism than to GABA neurotransmission. The speculation that GABAergic photoreceptors participate in the parallel ON-OFF pathways (responsive to increases or decreases in light intensity relative to the background) in the retina seems unlikely. There is excellent evidence, in various species, that the ON-OFF pathways are subserved by two classes of bipolar cell, second-order interneurons that make different types of

synaptic contact with the same photoreceptor^{9,11}. Thus, there is no need to invoke a presynaptic substrate for ON and OFF pathways.

The paper of Nishimura *et al.*² will certainly generate much debate. It addresses one of the most elusive issues of retinal cell biology — the identity of the photoreceptor transmitter(s). Although highly controversial, these findings appear to be based on solid technique and cannot be dismissed on technical grounds. Basic transmitter systems, such as GABA, appear well conserved in different species

and it is unlikely that the macaque monkey would be the only species with GABAergic photoreceptors. Before it can be concluded that some photoreceptors are indeed GABAergic, or even contain GABA and GAD, corroborative studies using other techniques need to be performed not only on other mammals but also on cold-blooded vertebrates, from which supporting *in vitro* data can more easily be obtained. □

Stephen Yazulla is in the Department of Neurobiology and Behavior, State University of New York, Stony Brook, New York 11794, USA.

Oceanography

Respiration down in the depths

from T. Roger S. Wilson

ALTHOUGH the deep ocean floor occupies a major part of the solid surface of our planet, it is still almost as alien to us as the surface of Mars. Recently, deep-ocean vehicles have been designed to sample and perform experiments on the sea floor. The results from the latest of these, reported in this issue by Clare Reimers *et al.* (*Nature* 320, 741; 1986), show that the oxygen consumption of the deep ocean floor at the sites sampled is surprisingly heterogeneous, even on the centimetre scale.

Most of what we know about the environment of the deep ocean has been obtained from samples and cores returned from abyssal depths to the surface. This sudden decompression, and other related insults, are disastrous to most organisms. Some of the bacterial population are apparently destroyed, although gross microbiological activity is often enhanced by decompression (Jannasch, H.W. & Wirsen, C.O. *Appl. envir. Microbiol.* 43, 1116; 1982). Even the inorganic chemistry of the sediment and pore-water is disturbed, so that from samples taken on shipboard it is impossible to determine accurately fundamental facts such as the oxygen demand of the sediments or the saturation of the pore-water with respect to carbonate minerals (Emerson, S. *et al. Earth planet. Sci. Lett.* 61, 220; 1982). It has therefore been necessary to develop *in situ* sampling and incubation techniques. These techniques are at an early stage of development — indeed, it is only a relatively short time since the first measurements of dissolved oxygen in deep-ocean sediments were made (Wilson, T.R.S. *Nature* 274, 354, 1978).

Reimers *et al.* now report a significant advance in the development of such devices, and give some very interesting preliminary results. Their seafloor lander has returned information on sub-seabed oxygen gradients measured directly within the upper layers of the sediment. Micro-electrodes which can give a spatial resolu-

tion of better than ten microns have been used in sediment studies for some time (N.P. Revsbech *et al. Limnol. Oceanogr.* 25, 403; 1980), but never before as part of a remote analytical module. Because of the problems of keeping a ship accurately on station several thousand metres above the landing site, the instrument is autonomous, discarding ballast to return to the sea surface automatically after sampling.

The vertical resolution of the new measurements (1 mm) and their precision in such a difficult environment is a technical *tour de force*. The results are fascinating, showing fine detail in the oxygen profiles, and significantly different profiles only a few centimetres apart. Horizontal heterogeneity is expected in near-shore sediments with their intense bioturbation, but few perhaps would have predicted it in the sediments of abyssal depths. We are probably not seeing the direct effect on pore-water oxygen of a recent stirring of the sediment because diffusion would quickly smooth out differences on this scale. What, then, is the mechanism?

Reimers *et al.* have a plausible suggestion for the heterogeneity that involves variability in microbial respiration caused by the redistribution of particulate organic carbon within the upper layers of the sediment. The high quality of these data confirms the value of the *in situ* technique. It is difficult to see how the insight given by these results could have been obtained in any other way.

Deep-ocean sediments hold a record of the climate of past ages and will, in the future, form the final sink for fossil fuel carbon dioxide. Even though the study of these sediments requires much ingenuity and patience, it would be unwise to ignore what the results can tell us. □

T. Roger S. Wilson is Principal Scientific Officer at the Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK.

- Altman, J. *Nature News and Views* 313, 264 (1986).
- Nishimura, Y., Schwartz, M.L. & Rakic, P. *Nature* 320, 753 (1986).
- Dowling, J.E. & Ripps, H. *Nature* 242, 101 (1973).
- Ishida, A.T., Kaneko, A. & Tachibana, M. *J. Physiol., Lond.* 348, 255 (1984).
- Lasater, E.M. & Dowling, J.E. *Proc. natn. Acad. Sci. U.S.A.* 79, 936 (1982).
- Yazulla, S. in *Progress in Retinal Research* (eds Osborne, N.N. & Chader, G.J.) 1 (Pergamon, Oxford, 1986).
- Miller, R.F. & Slaughter, M.M. in *Retinal Transmitters and Modulations: Models for the Brain* (ed. Morgan, W.W.) 123 (CRC, Boca Raton, Florida, 1985).
- Oertel, W.H., Schmechel, D.E., Tappaz, M.A. & Kopin, I.J. *Neuroscience* 6, 2689 (1981).
- Stell, W.K., Ishida, A.T. & Lightfoot, D.O. *Science* 198, 1269 (1977).
- Lasansky, A. *J. Physiol., Lond.* 285, 531 (1978).
- Nelson, R. & Kolb, H. *Vision Res.* 13, 1183 (1983).

The stability of zoological nomenclature

SIR—With the exponential growth of scientific publication, and the enormous broadening of the subject of biology, the majority of biological publications that use the names of animals have been written in the past 50 years, and the majority of authors are now not taxonomists. Taxonomy, however, remains one of the cornerstones of biology with the main practical function of providing animals with unequivocal labels. Instability in taxonomic nomenclature is highly detrimental to the study of biology, and the declared aim of the International Commission for Zoological Nomenclature is to promote stability for the names of animal species. We suggest that this organization has failed in its aim, and indeed that within the Code of Zoological Nomenclature, as presently constructed, lie sections that cannot be regarded as anything but mechanisms for promoting instability.

To take one example, the code states that a specific name should be of the same gender as the generic name, and if the species is subsequently removed to a different genus, the ending of the specific name must be changed to agree with the gender of the new generic name. Since many, perhaps most, newly described species eventually end up in a different genus, a great many such names will have to be changed. Indeed, since the chance of the new genus being of different gender is probably about the same as its being of the same gender, this provision of the code ensures that about half of the names of species will have to be changed if they move to a different genus.

Another practice that guarantees instability is that of resurrecting long-dead homonyms. If a newly published name turns out to be identical to an existing name (that is, both generic and specific names are the same) then it is clearly a homonym and must be changed to avoid confusion. This is entirely reasonable. But if the problem goes unnoticed at first and, meanwhile, one of the two species is moved to a new genus, we might reasonably suppose that the problem has solved itself, since the two species now have different generic names. Regrettably, this is not the case, since according to the code, and for no better reason than that both species used to have the same names, the name must still be changed in spite of the confusion that will be caused. History must be kept tidy.

It might be useful to give an actual example of this practice. There lives in Africa a fly whose larva, called the Congo floor maggot, sucks the blood of man: it is thus of medical importance. It was named *Musca luteola* by Fabricius in 1805. Nobody noticed at the time that in 1763,

Scopoli had already given the same name to a very different insect. In 1907, *Musca luteola* was moved to the genus *Auchmeromyia luteola* and there can be no confusion. Nevertheless, in 1980 the name was changed to *Auchmeromyia senegalensis* in the *Catalogue of the Diptera of the Afrotropical Region* because of the primary homonym and despite the fact that in the definitive works on this fly, including the medical literature, the name *Auchmeromyia luteola* is used throughout, and medical men have been informed that this is the only species of the genus of medical importance.

There are other devices in the code to ensure unnecessary name changes, for example the discovery of some ancient type-specimen, but nothing will be gained from an endless recital. Suffice to say that if the International Commission for Zoological Nomenclature expects to be taken seriously by the scientific community, it would do well to produce something better than the present document.

The response of taxonomists to this situation seems to be of two kinds. There are those (a minority, we think) who, having a legalistic turn of mind, do not question the code and fully approve of its rules. But far too many taxonomists wring their hands over the deplorable state of affairs, but are kept in place by the tyranny of the Commission: the Code must be obeyed. There is only one course open to us if sanity is to be restored and that is that some sections of the code must be consistently ignored.

The problems generated by strict adherence to the code are compounded by the current craze for cladistics (or phylogenetic systematics). Many taxonomists believe, almost as an act of faith, that a natural classification is there waiting to be discovered. We believe that classification is imposed by man, and furthermore it is usually subsequently 'discovered' to be wrong.

The changing of the classification with each and every publication on the phylogeny of the group is highly irresponsible. A particularly common and objectionable practice is the elevation of subfamilies to family rank. According to the code, the name of the original family has to be retained for the now restricted family, so that once again, the confusion is made compulsory under the code.

A better rule, if we really must split up families (a practice that seems to have little scientific merit), would be to forbid the original name to be used for the restricted family, in order that the otherwise inevitable confusion can be avoided.

To allow the present situation to continue simply devalues the family concept. However, there is no reason why the situation need arise at all. Phylogeny and

classification are different activities, and we see no reason why the current (and usually ephemeral) ideas on the one have to cause confusion to the other.

Y.Z. ERZINCIOGLU
D.M. UNWIN

Department of Zoology,
Downing Street,
Cambridge CB2 3EJ, UK

Interleukin terminology in disarray

SIR—Your comment on the unfortunate duplication of the term interleukin-4 (IL-4) for two distinct lymphokines is a very fair summary of the situation¹. We wish to explain how this confusion came about, and to point out that no effective framework exists for the naming of lymphokines.

When we first became aware of this duplication we informed Professor Honjo (21 December 1985), and established that we had priority of publication, as he had not received galley proofs. He initially agreed to change the name of BSF-1 to IL-5, but subsequently opted to duplicate our usage of IL-4. Our paper appeared on 17 January², and their's on 20 February 1986³. The situation was further confused on 7 March by the designation of a largely uncharacterized activity involved in amplifying the T-cell response as IL-4A⁴.

Unlike its biochemical counterpart, the nomenclature committee of the International Union of Immunological Societies (IUIS) has not been very active in this field (in fact the lymphokine subcommittee has never met). The term 'interleukin' was coined by an *ad hoc* group which defined IL-1 and IL-2⁵. Subsequent terminology has been determined by individual authors on the basis of priority of publication. Thus for example, the term IL-3 was proposed⁶, and has been widely adopted for the factor produced by WEHI-3 cells. It replaces about a dozen synonyms. Another *ad hoc* group attempted to introduce a unified nomenclature for factors active on B cells⁷, but this has not been universally accepted. These groups have usually arisen at scientific meetings, and have not necessarily represented all interested parties. It must also be pointed out that all these names have been based on biological activities, well before the protein sequences were established.

Our use of the term IL-4 was an attempt to clarify a complicated field. We had identified a lymphokine controlling eosinophil differentiation, and subsequently found it also had B-cell growth factor activity. We needed a name for this factor that was not linked to either biological activity, and we hoped to avoid the proliferation of names and consequent confusion that had occurred with IL-3. Because of our priority of publication we

feel justified in continuing to use the term IL-4 for this factor.

With the current burgeoning of interest in lymphokine research, this problem is likely to occur again. We therefore suggest that the IUIS should take appropriate steps to provide nomenclature guidelines for these molecules.

COLIN J. SANDERSON
GERRY G.B. KLAUS

National Institute for
Medical Research,
The Ridgeway,
Mill Hill,
London NW7 1AA, UK

1. Newmark, P. *Nature* **319**, 620 (1986).
2. Sanderson, C.J. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 437 (1986).
3. Noma, Y. *et al. Nature* **319**, 640 (1986).
4. Milanese, C., Richardson, N.E. & Reinherz, E.L. *Science* **231**, 1118 (1986).
5. Aarden, L.A. *et al. J. Immun.* **123**, 2928 (1979).
6. Lee, J.C., Keller, J. & Ihle, J.N. *J. Immun.* **128**, 2393 (1982).
7. Coutinho, A. *et al. Immun. Today* **4**, 332 (1983).

Overburdened African women

SIR—Within the context of the energetics of load carrying by African women^{1,2}, I wish to point to a condition which is known as Kikuyu bursa^{3,4}. This condition may provide a clue for some of the questions raised. As indicated in the article, Kikuyu women carry loads behind their backs supported by a strap. Compression of skin and subcutis between load and vertebral column and the horizontal—longitudinal vibrations of load during each step induces an anatomical structure in the subcutaneous tissue of the low-back area. This structure has the features of a classical bursa. This Kikuyu bursa partially absorbs the vibration forces related to load carrying.

The bursa, the histological structures of which cannot be distinguished from the prepatellar bursa (also known as housemaid's knee), may become enlarged and painful, and is often surgically removed. Extensive calcification may be present in the wall of this chronically inflamed bursa.

J.W. KOTEN

Weezenhof 2107,
65 36 JR Nijmegen,
Netherlands

1. Maloij, G.M.O., Heglund, N.C., Prager, L.M., Cavagna, G.A. & Taylor, C.R. *Nature* **319**, 668–669 (1986).
2. Alexander, R. McN. *Nature* **319**, 623 (1986).
3. Koten, J.W. *Ned. Tijdschr. Geneesk.* **111**, 2253 (1967).
4. Maathuis, J.B. & Koten, J.W. *Trop. geogr. Med.* **21**, 389–396 (1969).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a rather technical character which are not provoked by articles or letters previously published.

Discrepancies in length of myosin head

SIR—Mendelson¹ recently discussed discrepancies between various measurements of the length of the myosin head, this being an important parameter in any detailed description of the structural changes that cause muscle contraction. He criticized the view^{2,3} that recent data from crystals of myosin subfragment-1 (S1, equivalent to the myosin head) support the length of about 19 nm measured by electron microscopy^{4–7}, as opposed to the 12 nm value inferred from X-ray scattering measurements⁸. We contend that, on the contrary, the bulk of data, including those from S1 crystals, do indeed show the head to be about 19 nm long, and that there is evidence of insensitivity in the X-ray scattering method that could account for its lower estimate.

The myosin head is roughly the shape of an elongated comma, with the actin binding site located near its bulbous end⁹ and the tail part of the myosin molecule attached to its narrower (neck) end. The neck region of the head includes the regulatory light chain, and electron microscopy of shadowed myosin and S1 shows that intact heads (that is, containing the light chain) are approximately 19 nm long^{4,5} but apparently become shorter (the neck disappears) if the light chain is removed⁶. Most three-dimensional reconstructions of the actin–S1 complex have used S1 lacking this light chain, and they concur with the images of single molecules in showing a roughly 12 nm length^{7,9,10}. When the light chain is present, a longer length (at least 17.5 nm) is found, similar to the length of the intact head⁷.

In contrast with this picture, X-ray scattering data have been interpreted to show a length (maximum chord) of 12 nm (ref. 8), which does not change when the light chain is removed¹¹. The discrepancy between electron microscope and X-ray measurements is too great to be explained simply by the fact that the longer lengths quoted from electron microscopy are often contour lengths (as suggested by Mendelson¹). For example, even the maximum chord of S1 with the regulatory light chain present (measured by electron microscopy) is at least 15–16 nm (refs 7, 12), 3–4 nm greater than the X-ray scattering value. Taken alone, the failure of X-ray scattering to detect any change in length when the light chain is removed suggests that the light chain is located a distance from the centre of mass of the molecule equal to the measured radius of gyration (R_g), 3.5 nm (ref. 11). Such a conclusion is incompatible with the electron microscopy results, which show most of the light chain to be a long way from the centre of mass. A further inconsistency in the X-ray results is that R_g is unchanged when the head is attached to a substantial length of

the myosin tail¹². As Mendelson himself points out¹², this suggests that the region of the X-ray scattering spectrum that is examined to determine R_g could similarly be insensitive to the presence of an extended neck on the head, especially if the neck were flexible, with the result that the true length would be underestimated.

Crystalline S1 (ref. 3) contains the regulatory light chain and has an extended neck as judged by the barbed arrowheads formed when it binds to actin^{13,14}. Winkelmann *et al.*³ state that the length of the head is greater than 16 nm. The data from crystals are therefore consistent with the longer length seen in the electron microscope and do not agree with the short length.

Thus, measurements by all four different electron microscope approaches^{3–7} give striking agreement in their estimates of a head length of at least 16 nm. The shorter of these lengths are probably underestimates (because of averaging procedures and of difficulties in determining molecular boundaries — see refs 3, 7) and the true value is likely to be close to 19 nm, as measured in images of individual myosin molecules, where the length can be measured unambiguously from the tip of the head to its junction with the myosin tail. In our view, the disagreement between the conclusions from the electron microscope and the X-ray data point to a weakness in the X-ray scattering method in that it may fail to detect thin, probably flexible¹⁴, portions of protein molecules.

ROGER CRAIG

Department of Anatomy,
University of Massachusetts
Medical School,
Worcester,
Massachusetts 01605,
USA

JOHN TRINICK
PETER KNIGHT

Muscle Biology Division,
AFRC Institute of Food Research —
Bristol Laboratory,
Langford,
Bristol BS18 7DY,
UK

1. Mendelson, R. *Nature* **318**, 20 (1985).
2. Craig, R. *Nature* **316**, 16 (1985).
3. Winkelmann, D.A., Mekeel, H. & Rayment, I. *J. molec. Biol.* **181**, 487 (1985).
4. Elliott, A. & Offer, G. *J. molec. Biol.* **123**, 505 (1978).
5. Flicker, P.F., Wallmann, T. & Vibert, P. *J. molec. Biol.* **169**, 723 (1983).
6. Walker, M., Knight, P. & Trinick, J. *J. molec. Biol.* **184**, 535 (1985).
7. Vibert, P. & Craig, R. *J. molec. Biol.* **157**, 299 (1982).
8. Mendelson, R. & Kretzschmar, K.M. *Biochemistry* **19**, 4103 (1980).
9. Taylor, K.A. & Amos, L.A. *J. molec. Biol.* **147**, 297 (1981).
10. Toyoshima, C. & Wakabayashi, T. *J. Biochem. (Tokyo)* **97**, 219 (1985).
11. Mendelson, R. *Nature* **298**, 665 (1982).
12. Mendelson, R.A. & Wagner, P.D. *J. molec. Biol.* **177**, 153 (1984).
13. Rayment, I. & Winkelmann, D.A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4378 (1984).
14. Craig, R. *et al. J. molec. Biol.* **140**, 35 (1980).

Cortical convolutions

Colin Blakemore

Visual Cortex. Cerebral Cortex, Vol. 3. Edited by Alan Peters and Edward G. Jones. Plenum:1985. Pp.424. \$65, £54.93.

Models of the Visual Cortex. Edited by David Rose and Vernon G. Dobson. Wiley:1985. Pp.586. £49, \$77.95.

TWENTY years ago, when I was graduate student in Horace Barlow's laboratory at Berkeley, there were no more than half-a-dozen important papers about the visual cortex, all of them by David Hubel and Torsten Wiesel at Harvard Medical School. Hubel and Wiesel say that a crack in the glass of a projection slide they were using in their experiments led them to the discovery that each cell in the cat's primary visual cortex (area 17) responds to a line or edge at a particular orientation moving over its receptive field; that is, cortical cells are orientation selective.

That observation on its own was not remarkable. The idea that nerve cells in the visual system might act as detectors of specific and important *features* in the retinal image was very much in vogue. But in the hands of Hubel and Wiesel it was the key that unlocked a Pandora's box of treasures from which they pulled one jewel after another for a quarter of a century. Orientation columns, ocular dominance columns and their interrelated architecture, the uniformity of the modular structure of the visual cortex, the combination of inputs from the two eyes, the developmental plasticity of cortical cells, the sensitive period: all of these remarkable facts and phenomena yielded to their microelectrodes and, later, to their expertise in neuroanatomy. By 1966, Hubel and Wiesel had posed four major problems about the cat's visual cortex:

(1) How is the selectivity of cortical cells for orientation created from inputs (the axons of cells in the lateral geniculate nucleus) that are themselves responsive to light or dark spots, or to edges of any orientation?

(2) Can cells in the cortex be classified on the basis of their visual responses in a way that illuminates their functional organization, and relates to their morphology and their interconnection?

(3) How does this beautiful piece of neuronal machinery develop, and in what proportion is its maturation dependent on genetic and environmental factors?

(4) Just how do the properties of cells in the primary (striate) visual cortex, together with those in the neighbouring, additional visual regions of cortex, enable the animal to see the world?

Hubel and Wiesel had, even then, offered opinions on all of these issues, and if their ideas had been entirely correct I

guess that the whole thing would have been wrapped up long ago. It is an indication of the difficulty of these problems, and of Hubel and Wiesel's acuity in identifying them so long ago, that the four basic questions remain unanswered and still dominate this burgeoning field of research.

Now, four years after the award of the

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Trace of circuitry — a pyramidal neurone from layer III of the cat's visual cortex, which was injected with horseradish peroxidase after intracellular recording in an isolated slice of cortex maintained in vitro. The micrograph shows the cell body and some of the proximal dendrites, rich in spines, which are sites of synaptic input. The drawing shows the entire local dendritic tree, as well as the descending axon and its many collateral branches (thin lines). Scale bar = 50 µm. (Illustrations courtesy of Adrian Mason and Alan Larkman.)

Nobel prize to Hubel and Wiesel, there are literally thousands of researchers working on one aspect or another of the visual cortex. Instead of the half-dozen papers available in 1966, the poor graduate student who dares to venture into this overcrowded field is now faced with perhaps 20,000 publications. The increase in knowledge, insight and understanding, however, has not been proportional to the amount of new research, and the subject, though undeniably busy and exciting, is riddled with controversy and contradiction, not just because of rival hypotheses but, ominously, at the level of experimental data.

The two books reviewed here are intended, in their very different ways, to summarize how things now stand. That edited by Alan Peters and Edward Jones (both of whom are primarily anatomists) is the third volume in a series of five on the cerebral cortex. I suppose, then, that it has to be judged in conjunction with Vol. 1 (on the structure of the cortex) and Vol. 2 (on the general physiological properties of cortical neurones). Even so, it is surprisingly idiosyncratic and narrow in its coverage. Only the anatomy of the visual cortex gets thorough treatment, perhaps reflecting the interests of the editors as well as the fact that anatomy is very fashionable these days.

The high points for me were the gorgeous drawings of cortical neurones by Valverde (complete with doodles of the temple of Poseidon and of a playful kitten) and the scholarly review by Van Essen of the multitude of extrastriate visual areas described in the owl monkey by Allman and Kaas, and in the macaque by Zeki and others. The physiology of the primary visual cortex is largely left to Henry, who describes the flow of signals through the cortical network in terms of parallel "streams" based on input from the well-established functional groups of retinal ganglion cells (X, Y and W cells) in the cat. Henry also tackles (but does not answer) the long-standing question of whether cortical cells should be seen as highly specific detectors of information-rich features (edges, directional motion, corners and so on) or whether they might act as spatial filters, providing an essentially (behaviourally) neutral but (informationally) efficient description of the visual scene in terms of its component spatial frequencies and their phases. The latter approach (which grew out of Campbell and Robson's psychophysical evidence for spatial frequency channels underlying the contrast sensitivity of human observers) interprets orientation selectivity as a mechanism for limiting to a single dimension the spatial analysis performed by each module of cortical cells.

The second book, edited by David Rose (a neurophysiologist and psychophysicist)

and Vernon Dobson (a modeller) is utterly different, even though it shares some contributors (Peters, Henry, Braitenberg) with the first. Rose and Dobson had the bright idea of inviting, as they say, "75 of the most prominent theorists in the field" to write brief and speculative articles on their specialities. Miraculously, most of them not only accepted but met the tight deadline for submission. The resulting 57 chapters must have been an editorial nightmare, but Rose and Dobson have done a remarkable job of bringing some degree of order to potential chaos. The consistency of style and extent of cross-referencing are impressive. The glossary and "associative index" at the end are very useful. Like the Peters and Jones volume, however, this book is not really the answer to the naive student's plea for a pilot to guide him or her between the rocks and into a safe harbour of understanding about the visual cortex. Instead (to pursue this metaphor) it is a motley collection of brigands, mermaids and purveyors of unproven aids to navigation, who force or tempt or trick the reader towards their favourite corners of the coastline. To the aficionado, steaming a familiar course through these busy waters, it is, however, a valuable book, capturing the diversity and excitement of this field as well as its complexity, difficulty and variation in quality. What a shame, though, that there is no chapter from Hubel, Wiesel or any of their close collaborators.

As the title implies, the book is especially concerned with hypotheses about the cortex, though there is a mass of experimental data in it as well. The enforced brevity has produced a rather telegraphic style, but many of the contributors have managed to speculate and to articulate their worries in a manner rarely seen outside the bar at conferences.

Within this diverse collection it is possible to discern some important general themes. First, there is the notion of the cortex as an associative system, exploiting the statistical properties of the visual image to detect functionally important coincidences both during development and when finally mature.

Second, there is the recognition that inhibition (perhaps even modifiable inhibition) within the cortical circuit, involving the transmitter GABA, is crucially important for creating the selectivity of cells. Sillito's observation that orientation selectivity is sometimes virtually eliminated during the iontophoretic application of the GABA antagonist bicuculline obviously casts doubt on Hubel and Wiesel's original model in which this property was thought to be created by converging excitatory inputs from geniculate axons with their receptive fields aligned in a row. Bonds and DeBruyn provide a succinct review of this problem, and are forced to the conclusion that a cooperative network

controlled by inhibition is unlikely to account for the entire orientation columnar structure of the cortex without "some *intrinsically specified* neurons that rely on mechanisms other than network inhibition". I agree, and I suspect that the clusters of innately specified "simple" cells found even in visually inexperienced animals, perhaps dependent on crude orientation selectivity seen at retinal or geniculate levels, may provide the clue to how the orientation columns are laid down.

The third and fourth themes are subjects that Hubel and Wiesel themselves (now working separately) are pursuing, namely the way in which information flows through the architecture of the striate cortex (and on to extrastriate areas), and the relation of structure to function at the level of individual neurones. The *outputs* of the cortex are distinctly segregated both vertically, into different layers, and horizontally, into blobs, stripes or clusters of cells. That discovery simplifies a little the daunting problem of understanding the circuitry of the cortex, which is being revealed in all its confusing glory by various techniques including intracellular injection of horseradish peroxidase. What seems to be missing here is information about the relative *effectiveness* of the different synaptic connections that are seen: there seems to be a mismatch between what we know of the inputs to cells from their physiology and what the anatomists say they receive.

Then, there are the real modellers — those who use mathematics or computation to tell us how the cortex ought to be. Kenneth Craik and David Marr, both of whom died in their thirties, had a huge influence on this field, teaching us that the brain itself builds a model of the world and that the world sets certain formal restrictions on the computational operations needed to interpret the retinal image.

Finally, Rose and Dobson themselves, following the views of the philosopher of science Nicholas Maxwell, criticize "pure empiricism" and argue that the sort of controversy and contradiction that this field has seen could be avoided if there were prior agreement on a systematic and exhaustive exposition of all possible hypotheses. I must say that I'm sceptical. Twenty-twenty hindsight is a wonderful aid to the design of elegant models, and a model based on known facts can still be good if it has the power to predict future discoveries. If they are saying that we ought to *think* more about our research, then I would agree with them. But major discoveries in science have a way of depending as much on cracks in projection slides as on *a priori* insight. □

Colin Blakemore is Waynflete Professor in the University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK.

Plants and planet back in time

Peter R. Crane

Geological Factors and the Evolution of Plants. Edited by Bruce H. Tiffney. Yale University Press: 1985. Pp. 294. \$25, £25.

ON AN ecological timescale the complexity of interactions between plants and their environment is widely recognized, but in geological time the nature of these interactions and the extent to which they may influence patterns of environmental or evolutionary change have received little attention. This symposium volume is therefore a notable contribution. It covers a variety of plant-environment interactions and raises important questions that cut across disciplinary boundaries. Three chapters deal with the relationship between continental configuration, models of atmospheric and oceanic circulation, and phytogeographic patterns in the Devonian and early Carboniferous, two with the early development of continental ecosystems, while the remainder are concerned with the origin of autotrophic eukaryotes, the biological and physical processes associated with wildfire, and the relation between ecology and evolution in Pennsylvanian coal-swamps.

Beerbower's essay on the development of continental ecosystems is a stimulating synthesis of palaeontological, sedimentological and physical data. He outlines a "non-determinate" model for the origin of a terrestrial biota by arguing that critical, and possibly energetically expensive, adaptations for terrestrial life would have been subject to "counterselection" in dominantly resource-poor, intensely disturbed, continental environments prior to plant colonization. In contrast, Chapman's perspective on the geological and biochemical factors relevant to the origin of land plants is openly determinist. Re-stating the hypothesis developed by Lowry, Lee and Héban, he suggests that coumaric acid and other products of phenylpropanoid metabolism may have functioned as screens of ultraviolet radiation in land plants before the development of an effective ozone layer. Subsequently, increased photosynthesis and high oxygen levels would have precluded the need for internal ultraviolet shields and permitted the development of the oxygen-dependent pathways of phenylpropanoid metabolism involved in the biosynthesis of lignin and cutin. These are the critical compounds involved in support, water conduction and waterproofing in land plants.

Predictably, neither perspective is wholly satisfactory. Chapman's suffers (as

he acknowledges) from the uncertainties in our understanding of oxygen levels in the late Proterozoic and early Phanerozoic, while the tests that Beerbower offers of his null hypothesis are weakened by the extent to which corroboration or refutation is critically dependent on a nexus of subsidiary hypotheses. Clearly, there is considerable scope for carefully directed neobotanical and palaeobotanical research in this area.

In addition to Beerbower's contribution, the other paper that most directly addresses the historical dynamics of plant-environment interactions is that by DiMichele, Phillips and Peppers. These authors build upon a series of earlier papers, and characterize Pennsylvanian coal swamps as edaphic islands with low diversity, rather uniform communities that remained discrete from other lowland

vegetation for long periods; vegetational changes resulting from increases or decreases in abundance, phylogenetic changes in swamp plants or new introductions from other lowland environments were concentrated during intervals of stress or disturbance. Their analysis is unique among studies of pre-Cenozoic plant communities in the level of detail available, the extent to which macrofossil and palynological data have been integrated, and in the convincing links established between large-scale ecological and phylogenetic patterns. In all these respects it is a model study that other researchers, pursuing the themes of this successful symposium volume, would do well to emulate. □

Peter R. Crane is Associate Curator of Paleobotany in the Field Museum of Natural History, Chicago, Illinois 60605, USA.

Isotopic man

Colin A. Russell

Frederick Soddy (1877–1956). Edited by George B. Kauffman. *Reidel*:1986. Pp. 239. Dfl. 130, \$44.50, £36.25.

WHY should nuclear chemistry ever be called nuclear physics? Today there is plenty of room for each of them as well-differentiated entities. But it was not always so. This was a severe problem that greatly vexed Frederick Soddy, a man whose research laid the foundation of both sciences.

Soddy has been well-known to generations of students through two remarkable works, *The Chemistry of the Radioelements* (published between 1911 and 1914) and *Isotopy* (written with J.A. Cranston and published in 1954). In the 40 years between these publications he received the 1921 Nobel Prize for Chemistry, and was for 17 years Dr Lee's Professor of Chemistry at the University of Oxford. His fame rests on his "contribution to our knowledge of the chemistry of radioactive substances and his investigations into the origins of nature of isotopes", as the Nobel citation neatly puts it.

In fact it is to Soddy, more than anyone else, that we are indebted for the concept of an "isotope". It is therefore quite astonishing that so distinguished a founder of the nuclear sciences should have been bothered by the semantic problem mentioned above, and still more that in the four decades from 1914 to 1954 he should have done relatively little scientific work. To be sure, there continued to flow a stream of books, lectures and papers, but these were not on nuclear chemistry (or nuclear physics) but the remote subjects of sociology and economics. That this should be the case makes the subject of

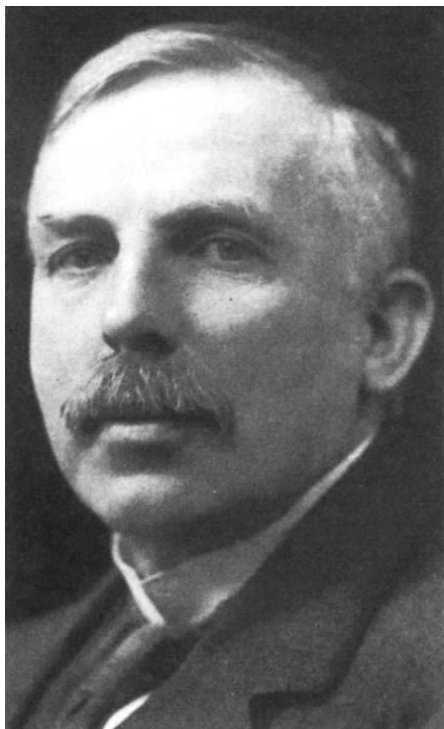
this book a scientist of more than ordinary interest. For if the history of science is to be worth anything at all, it must address itself to the question of how scientists tick, have ticked (and presumptively will tick). On that account alone their "failures" are as important as their "successes". What is particularly sad about Soddy is that, unlike other scientists such as Polanyi who have changed course in mid-stream, he made almost no impact in his new sphere of activity.

With Frederick Soddy it is all too easy to come to quick conclusions. He peaked early; he was hampered and indeed affronted by the First World War (in which his contemporary Henry Moseley was killed); he had an awkward personality; he did not suffer fools gladly; and he resented intrusion into his own field. Yet none of these judgements is sufficient explanation for the almost bizarre story of a man who failed not only to fulfil his early promise in chemistry, but also to impress his views on a world which most chemists fear to enter, that of the social sciences.

It would have been splendid if this book took us further and gave us a complete picture of the man and his work, together with a sustained analysis of his life. Regrettably it does not do so, nor could this reasonably be expected in a collection of essays from a dozen authors with minimum editorial correlation. So meticulously has the editor preserved the texts of his contributors that he has not even "Americanized the spelling[]" of British authors". Nor, it may be added, has he unified the style of references or notes, not over-important in itself but symptomatic of a degree of literary *laissez faire* that makes an integrated treatment impossible. Predictably, there is plenty of repetition — I lost count of the references to radioactive decay series with which Soddy was so closely associated.

We can, however, be grateful for the

assembly between two covers of so much material on one person. Three important papers by Soddy are reprinted, including his letter to *Nature* of 4 December 1913 in which he introduced the term "isotopes". One of the most convincing chapters is that by A.D. Cruikshank, dealing with Soddy's long years at Oxford, while the account of the reception of his ideas in Japan and the USSR will be of great interest to anyone concerned with the wider development of sciences in those countries. His extreme popularity in the USSR



Frederick Soddy — a founder of both nuclear chemistry and nuclear physics.

is all the more remarkable in the light of the anti-Marxist emphases in his works on economics.

Perhaps one of the most provocative notes is sounded by L. Badash in his chapter "The Suicidal Success of Radiochemistry". So far as Soddy was concerned there was little more to do after 1914 except tidying up. He had no wish to join the lemmings. For radiochemistry the war could hardly have come at a worse time, with confirmation of isotopes announced by four laboratories in four countries within one month of 1914. Soddy was not the only scientist to find the discontinuities of war sufficient cause for substantial disengagement from chemical activity; Percy Frankland was another. But it seems that amongst the myriad casualties of that conflict we must also number the scientific creativity of Frederick Soddy. From a near oblivion for which he was in part deliberately responsible this book may rescue him in the nick of time. □

Colin A. Russell is Professor of History of Science and Technology at the Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.

Chemical warfare against viruses

B.W.J. Mahy

Approaches to Antiviral Agents. Edited by Michael R. Harnden. Macmillan, London/VCH, Deerfield Beach, Florida: 1985. Pp.326. £45, \$72.50.

BECAUSE viruses are intracellular parasites, drugs which block virus replication usually damage the host cell, with consequent harmful side-effects. For this reason (and in contrast to the many antibiotics available to control bacterial infections), only a handful of chemotherapeutic agents have been identified for use against human virus diseases; currently, five such drugs are licensed for use in Britain, four against herpes and one against influenza. Prophylactic immunization has thus been the main weapon in the war against virus infection, but some infections are difficult to control in this way and interest in antiviral agents has continued.

This multi-author volume considers the mechanisms of action, biological properties and clinical application of known antiviral agents, and the available approaches to the discovery of new compounds. After a brief assessment of the world-wide clinical need for antiviral drugs, five contributions are devoted to specific inhibitors of viral replication. The recent discovery of a range of nucleoside analogues active against herpes infections has given a considerable boost to this area of research. These compounds do not harm uninfected cells but become activated in herpes-infected cells owing to specific phosphorylation by virus-encoded enzymes, with the result that virus-infected cells alone are destroyed. Such compounds might not have been discovered without a detailed knowledge of molecular events in the virus replication cycle, so other potential targets for selective inhibition are considered here. Two further chapters review published work on synthetic nucleosides, and must include descriptions of every analogue which has ever been tested against a virus, while the traditional random screening of synthetic or natural compounds with potential antiviral activity is described in detail in another two chapters.

There follow four accounts of the modulation of host defence mechanisms, which mostly discuss interferon and related substances. Interferon is one of the body's natural defences against viruses but, despite recent advances made through the application of recombinant DNA technology, the practical uses of interferons, either as antivirals directly or as immunostimulating agents, are at pre-

sent very limited. The final chapter deals with the clinical evaluation of antiviral drugs.

On the whole, the eleven contributions are thorough, well-referenced and contain much stimulating discussion. There is a mood of optimism throughout, implying that new antiviral agents are just around the corner and that, given sufficient research effort, new drugs can be developed and put into clinical use. Like bacteria, however, viruses readily mutate to produce resistant strains in the face of inhibitory compounds, and perhaps too little attention is paid to such problems. Nevertheless, research workers involved in chemical warfare against viruses should have access to a copy of this book, and there is much of interest here for the general virologist as well. □

B.W.J. Mahy is Director of the Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK.

Parasite parts

Keith Vickerman

The Biology of Trypanosomes. Edited by Leslie Hudson. Springer-Verlag: 1985. Pp. 183. DM 124.

SYMPOSIUM volumes on trypanosomes are now appearing at the rate of three or four a year. Most of them give prominence to how the parasite artfully eludes the host's immune attack or how the host over-reacts to the parasite's presence. This collection of nine short essays is no exception. It is not, as the publisher claims, "the only recent review of trypanosome biology as a generic entity". It deals with only limited aspects of experimental work on the two species that infect man, *Trypanosoma brucei* and *T. cruzi*, and is thus far more restricted than (for example) the recent number of the *British Medical Bulletin* devoted to trypanosomiasis.

What the various contributors do bring out nicely is how the host — parasite relat-

ionship is reflected in the nature of the surface of these two very different parasites. The sleeping sickness trypanosome, *T. brucei*, multiplies in body fluids and escapes host antibody assault by repeatedly changing the exposed antigen in its monomolecular glycoprotein coat; in the book, the chemical basis of this process of antigenic variation is discussed in detail. The parasite of Chagas' disease, *T. cruzi*, multiplies inside macrophage or muscle cells and has no need for antigenic variation. The stage in the life cycle that ventures out into the host's blood to infect new cells carries on its surface a variety of glycoproteins, and evidence is presented here that many of these participate in entry into the host cell by binding to receptors or enzymatically exposing receptors for the parasite.

Two of the articles are strikingly novel. One introduces the intriguing possibility that DNA sequence rearrangements, known to play a part in antigenic variation in *T. brucei*, may be more widely involved in trypanosome differentiation. The other is a spirited attack on the problem of how non-immune mechanisms limit parasite numbers in the blood in *T. brucei* infections.

Some of the most exciting areas of current trypanosome research are conspicuously absent from the book; indeed the most flourishing enterprise of all — the molecular genetics of antigenic variation — was purposely excluded by the editor on the grounds that it is amply covered elsewhere. Yet half the topics in the book have also appeared in other reviews of late, especially macrophage-mediated immunosuppression in *T. brucei* infections and autoimmunity in Chagas' disease. The editor also regrets omission of an article on the sex life of trypanosomes. So do we all. For too long this has been the subject of rumour and innuendo, and it is high time that somebody put a stop to it by setting down the incontrovertible evidence in black and white. □

Keith Vickerman is Regius Professor of Zoology, University of Glasgow, Glasgow G12 8QQ, UK.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Spermatozoa of the diurnal sand rat attached to the zona pellucida of the laboratory rat, a heterologous species. The picture is taken from Gamete Surfaces and Their Interactions, edited by James K. Koehler and published by Alan R. Liss. Price is \$40, £37.

Polish science in crisis

This article is from a text prepared by the "Social Committee for Learning" (SKN), an unofficial organization of scholars established in Poland after the imposition of martial law. The membership of SKN is secret, but is believed to include some leading members of the Polish Academy of Science.

UNDERFINANCING of science and learning in Poland is a chronic condition and not just a result of the most recent crisis. In the 1970s, when ministerial institutes and research and development centres were being built at an alarming pace, and when colleges were multiplying in places with no tradition or base of trained personnel, Polish investment in science and learning was the lowest of all the Comecon countries, amounting to barely 2.2 per cent of the distributed national income in 1979.

Even then, the danger of the situation was pointed out, although in cautious terms, by various committees of the Polish Academy of Sciences (Poland 2000, Committee of the Science of Science). After August 1980, the crisis in science and learning was both institutional and material, and was widely and vociferously discussed, *inter alia*, in the fifty-second session of the General Assembly of the Polish Academy of Sciences. In November 1981, the Presidium of the academy proposed a fundamental reorientation of financial support, intended to prevent the massive run-down of science and learning, and indicated the minimum investment required even to preserve the status quo. But none of these warnings affected the policy.

In 1983, investment in science in Poland had fallen below one per cent to 0.8 per cent of the national income compared with 4 per cent in Czechoslovakia, 3.6 per cent in Hungary, 4 per cent in East Germany and 4.6 per cent in the Soviet Union. In 1984, spending was apparently up to 1.2 per cent. According to the calculations of the Science Policy Institute, this investment works out today in Poland as \$25–28 per inhabitant, compared with in Bulgaria \$75–80, Czechoslovakia \$150–160, East Germany \$170–180, the Soviet Union \$140–150 and, in the industrialized Western countries, \$250–500. In this respect, Poland must count itself among the backward countries of the Third World.

Out of the total funding, the amount allotted in Poland to basic research, of fundamental significance to national culture and for the future of the developing applied sciences, is 14 per cent (whereas it should be 30–40 per cent); 85 per cent goes to technical sciences and technology in the broad sense. The bulk of these slender means thus goes on work which has little to do with academic research, *sensu strictu*.

The value of fixed assets in scientific establishments amounted in 1983 to around 305,000 million zloty, but depreciation in the higher education sector accounts for 28 per cent of these assets and, in the case of equipment, 78 per cent (for the academy, 75 per cent). The value of equipment, after allowing for depreciation, was 10 per cent of all fixed assets.

The value of apparatus imported from the capitalist countries over the past five years amounts to barely 4 per cent of the total investment [in this sector]. This has meant that more and more imports are "unique", and that there is little access to up-to-date equipment. Spare parts are difficult to obtain, not only because of a lack of hard currency, but also because the apparatus itself is obsolete. The state of research laboratories often makes it impossible to satisfy the criteria of precision demanded by reputable scientific journals, which leads to difficulty in getting work published.

Cooperation with other Eastern-bloc countries, publicized as a remedy for shortages of equipment, is [but] empty words, because the necessary instruments are not produced in the bloc for lack of access to Western technology.

Outsiders

The most serious consequence of all these problems is the rapid loss of the possibility of cooperation, and even of a common language between young Polish scientists and their foreign colleagues; at the same time, Polish science is becoming more parochial. Within a few years, a young worker trained in Poland by a prominent scholar, but with archaic apparatus, insufficient literature and decreasing contacts, will become an "outsider" [the English word is used], of no interest, even as a working hand. Young PhDs now working in the experimental sciences may be the last generation for many years to be trained to a standard comparable with world level.

Access to the scientific literature is an integral part of the research laboratory. Here, too, the situation is very bad. Allocations for buying books and journals have been falling since the 1970s. In 1980, the Polish Academy of Sciences received 60 per cent of the allocation for 1974, and in 1983 barely 15 per cent of that sum. Some institutes have now virtually given up buying books in order to maintain con-

tinuous runs of the journals to which they subscribe. Because of price rises, the number of subscriptions has nevertheless decreased, while it is impossible to take out subscriptions to new titles, which increase every year by between 10 and 15 per cent.

The pay situation of scientific workers has been systematically deteriorating. Real pay in the science sector has decreased the most in comparison with other sectors of the economy, namely by 32.3 per cent, and for professors and assistant professors by 50 per cent. This state of affairs has led to a rapid drain of trained personnel from science. In 1981–83, some 50,000 people left, equivalent to an annual drop in the work-force of around 8 per cent. The documents of the nineteenth Plenum of the Central Committee of the Polish United Workers' Party indicate that between 1980 and 1983, employment in the development units of production enterprises fell by 32 per cent.

The drain of highly qualified engineering and technical personnel paralysed the work of some institutes. In universities and colleges the losses were less dramatic, but it must be remembered that the production of a new generation of scholars requires a steady influx of young personnel. Those who leave are often the most valuable and energetic, who seek other ways of realizing their aspirations, often by going abroad.

The Nineteenth Plenum of the Central Committee of the Polish United Workers' Party also announced increases of investment in science of 2 per cent in 1986 and 3 per cent in 1990. Even assuming that these promises are kept, these steps are not enough to halt the run-down of science. In five years, things will be worse than today even if the process of decline is partly retarded.

To cure the paralysis of science, it would have been necessary, even in 1985, to assign approximately 40,000 million zloty in 1983 prices for instruments and materials, rising to an annual 77,000 million by 1995. A hard-currency grant on the scale of several hundred million US dollars would also have been necessary. According to calculations by the Institute of Science Policy, investment in science would have to increase by 22 per cent a year simply to offset the wear and tear on equipment and buildings. Even to prevent further deterioration would require in-

vestment to be increased by 9.3 per cent annually.

To be sure, the salaries of scientific workers are being increased, especially those of professors and persons holding directors' posts, and in the eyes of the government this, together with administrative pressure, constitutes one of the means of "normalizing" the academic community. The principle of this "normalization" is simply: "Self-government for the normalized, in other words for those who do not wish to be self-governing." Leaving aside the pious wishes and slogans, that is what the political meaning of both the political catchphrases and the real intentions of the government boil down to.

The material conditions referred to above are reflected in the process of educating students. There is a shortage of rooms, libraries, apparatus and literature. Access to copiers, which could partly compensate for the lack of literature, is restricted for political reasons. At the same time, there is increasing interference by the ministry in the curricula which, hitherto, have been the responsibility of the relevant schools and courses. While classes for courses in the humanities have been limited to a maximum of 3,200 hours and in the natural sciences to 4,200 hours, the number of obligatory classes of a socio-political and military nature is constantly increasing. No more than 60 per cent of the time is left for course-related classes, which means that the standard of graduates will decline both from the point of view of preparation for a job and as candidates for future research.

For some years, there has been a tendency world-wide to extend the studies that prepare a fraction of the graduates of universities and higher colleges for future academic and creative work. In Poland, on the other hand, all doctoral studies in higher educational institutions were closed down in 1982. (There is now talk of their being restored in certain fields.)

The only possibility of obtaining a PhD is by working as an assistant lecturer. But since the number of such posts is linked to the limits on student intake, the influx of new assistant lecturers is very small, particularly in large universities and colleges. (In the academy, where in some centres doctoral studies still exist, intake is governed by political criteria.) The outcome will soon be a generation gap which, as is known from experience after the Second World War, is very hard to fill.

Inverted

Stipends for doctoral research will not provide a remedy, because there will simply be no assistant lecturers in the universities and colleges. In some institutions and departments (including the Polish Academy of Sciences) the service and age pyramid has already been completely in-

verted: there are several assistant professors and professors for every assistant lecturer, while in some laboratories only professors and assistant professors are employed, and they have nobody to teach.

The whole system of training personnel in Poland is choked: PhDs trained in higher education institutions and in the institutes of the Polish Academy of Sciences are not employed either in the institutes of the production ministries or in development centres or in general education, or in production enterprises. The cause is a personnel policy which prevents the rotation of workers which is essential in any research institution.

The composition of personnel in the specialized academy centres is likewise very uneven. No attempt is made to rectify this by an appropriate financial, investment and housing policy which would induce highly qualified scientific workers to seek work in the weaker centres and induce these centres to employ them.

Politics takes precedence over appraisal on merit in decisions about the establishment of new universities. The example of the creation of the University of Szczecin evokes disquiet: the lack of accommodation (requisitioned from secondary schools in spite of protests from the general education sector) and lack of personnel (which does not prevent political purges) do not augur well for the development of centres created in this way and in these conditions.

Scientific advance and the development of high-level personnel is much hampered by the material and pay conditions we have described. Most assistant lecturers and young PhDs are compelled to supplement their income, because the salaries of these posts are among the lowest. The impediments to foreign travel become greater all the time, which is especially vexatious in fields where it is virtually impossible to carry out reasonable work in Poland for a *Doctor Habilitatus* degree.

At a time that is so difficult for science, instead of helping to raise the level of trained personnel in ways that do not demand expenditure, and that would unblock the choked system, the government — again for political reasons — puts difficulties in the way of trips abroad on stipends and scholarships. As a result, the choice of topics for the degrees of PhD and *Doctor Habilitatus* are increasingly dictated not by the needs of society, but by the feasibility of carrying out the work under existing circumstances. In addition, it puts difficulties in the way of access to scientific work for graduates for whom there are no jobs.

As far as the development of research is concerned, the chief impediment at present is a lack of material resources. But for trained personnel, the main difficulties lie in the lack of thought for the future, a narrowly pragmatic approach to science

and political restrictions. Maintaining the growth of trained personnel at an acceptable level would be possible with a relatively small financial outlay and a sensible policy (and, of course, with an end to the rule that every associate professor or PhD must find employment in a higher educational institution or research centre). This would reduce the consequences of the present crisis for the future of science and national culture.

Policy

All the expert studies carried out over the past few years by competent teams of scholars have suggested the adoption of a policy which is the opposite of that at present in force; specifically, the establishment of priority directions for basic and long-term research, particularly for the health service, matters related to food and the protection of the environment; education and culture; the continuation of basic research in fields in which Polish science has an established place in the world; the maintenance of permanent contacts with science and technology throughout the world; and, even if at present one cannot afford many projects, halting the run-down of the material base and the "bloodletting" of personnel.

Undoubtedly a key problem for links between science and the economy and culture is the present organization of the economy in which industrial plants are not interested in making use of new solutions. The so-called economic reform of the 1980s has brought no improvement. So long as the production sector shows no interest in technical progress or in the quality and competitiveness of its products, financial and administrative pressures and reorganizations in the research sector will be unable to yield any improvement to the economy.

A clear distinction must be made between basic research, which is necessary for culture, environmental protection, health and so on, and which must be financed out of the state budget, and short-term applied research, financed and controlled directly by customers who want to, or who can, utilize the results. The present reorganization of the base of the scientific economy, which consists of an excessively developed network of centres, cannot carry out such a role, even if the centres were to function efficiently, so long as they have no really interested customers.

In research financed from the state budget, a certain number of centres with good equipment and highly qualified personnel should undoubtedly operate. It would be essential, however, to make their functioning independent of departments and low-level administrators, whose personal ambitions and interests have a deleterious influence on their work and to make it possible for them to do research unhampered by political factors. □

cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain and its expression in tumour cell lines

Christer Betsholtz*, Ann Johnsson†, Carl-Henrik Heldin†**, Bengt Westermark*, Peter Lind†**, Mickey S. Urdea§, Roger Eddy||, Thomas B. Shows||, Karen Philpott¶, Andrew L. Mellor¶, Timothy J. Knott* & James Scott*

* Department of Pathology, University Hospital, S-751 85 Uppsala, Sweden

† Department of Medical and Physiological Chemistry, Box 575, Biomedicum, S-751 23 Uppsala, Sweden

‡ Wallenberg Laboratory, University of Uppsala, Box 562, S-751 22 Uppsala, Sweden

§ Chiron Corporation, 4560 Horton Street, Emeryville, California 94608, USA

|| Roswell Park Memorial Institute, Department of Human Genetics, Buffalo, New York 14263, USA

¶ Transplantation Biology Division and # Molecular Medicine Research Group, MRC Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

The amino-acid sequence of the precursor of the human tumour cell line-derived platelet-derived growth factor (PDGF) A-chain has been deduced from complementary DNA clones and the gene localized to chromosome 7. The protein shows extensive homology to the PDGF B-chain precursor. Expression of the PDGF A-chain gene is independent of that of the PDGF B-chain in a number of human tumour cell lines, and secretion of a PDGF-like growth factor of relative molecular mass 31,000 correlates with expression of A- but not B-chain messenger RNA.

HUMAN platelet-derived growth factor (PDGF) consists of dimers of homologous polypeptide chains, denoted A and B (refs 1, 2). Whether PDGF is a heterodimer or a mixture of homodimers is not known, but the dimer structure is functionally important, since reduction irreversibly destroys the biological activity of PDGF. Connective tissue-derived cells display high-affinity cell-surface receptors for PDGF and respond to PDGF by receptor autophosphorylation, tyrosine phosphorylation of cytoplasmic substrates, increased cytoplasmic calcium concentration, activation of protein kinase C, cytoplasmic alkalization, reorganization of actin filaments, specific gene expression and DNA synthesis (reviewed in ref. 3).

The B-chain precursor is encoded by the *c-sis* gene, the cellular counterpart to the transforming gene *v-sis* of simian sarcoma virus (SSV)^{2,4-6}. The human *c-sis* gene has been mapped to the long arm of chromosome 22 (ref. 7) and has been shown to be transcribed in several human tumour cell lines⁸⁻¹¹ as well as in certain normal cell types such as endothelial cells¹², placental cytotrophoblasts¹³ and activated macrophages^{14,15}.

The primary translation product of the *v-sis* gene undergoes dimerization and proteolytic processing at the N- and C-terminals, yielding a product of relative molecular mass (M_r) 24,000 (24K) which resembles a dimer of PDGF B-chains and is recognized by anti-PDGF antibodies¹⁶. There is ample evidence that SSV-induced transformation is mediated by a PDGF-like growth factor. First, SSV-transformed cells contain and release a PDGF agonist activity¹⁷⁻²². Second, acutely SSV-transformed human fibroblasts are morphologically indistinguishable from PDGF-stimulated cells, and more significantly, their transformed phenotype is reverted by the addition of anti-PDGF antibodies to the culture medium²³. Studies of the transforming protein of SSV have indicated that assembled PDGF B-chains alone form an active mitogen. Furthermore, amino-acid sequence analysis of porcine PDGF has revealed that this dimeric factor contains only one type of chain, corresponding to the human B-chain²⁴.

Evidence that homodimers of PDGF A-chains also have biological activity was recently obtained from studies of a PDGF-like mitogen released from a human osteosarcoma cell line, U-2 OS. This factor²⁵, which binds to the PDGF receptor, was found to be a homodimer of a polypeptide chain that displays a chemical fragmentation pattern, chromatographic behaviour and N-terminal amino-acid sequence identical to that of the PDGF A-chain²⁶.

We report here the complete primary structure of the PDGF A-chain precursor deduced from its complementary DNA sequence, its structural relation to the PDGF B-chain precursor, the chromosomal localization of the gene and its expression in human tumour cell lines. We also present data showing that the release of biologically active 31K PDGF-like growth factors by human tumour cell lines correlates with PDGF A-chain but not B-chain gene expression.

PDGF A-chain cDNA

A λ gt10 cDNA library was constructed using poly(A)⁺ RNA purified from the human clonal glioma cell line U-343 MGaC12:6. This cell line was chosen because it produces higher quantities of PDGF receptor competing activity than do other cell lines investigated. An 86-base-pair (bp) oligonucleotide probe (PDGF-A-1) corresponding to the N-terminus of the PDGF A-chain amino-acid sequence (Fig. 1) was synthesized and used to screen the library (2×10^6 recombinant clones) at low stringency. Of 48 positive clones, 4 hybridized to a 37-bp oligonucleotide probe (PDGF-A-2) directed against a mid-portion of the A-chain amino-acid sequence and were selected for further analysis.

DNA sequence analysis showed that the four clones overlapped and contained inserts of 800-1,400 bp (not shown). The complete nucleotide sequence, determined from one clone (D1), is shown in Fig. 1. The longest open reading frame of this 1.3-kilobase (kb) cDNA predicts a PDGF A-chain precursor protein of 211 amino acids ($M_r \sim 23,000$), and an in-frame termination codon is situated 81 bp upstream of the putative translation initiation site. Two additional ATG triplets lie within the 387 bp of the 5'-untranslated region sequenced, but these do not conform to the consensus for translation initiation²⁷ and predict only short polypeptides.

** Present addresses: Ludwig Institute for Cancer Research, Biomedicum, S-751 23 Uppsala, Sweden (C.-H.H.); KabiGen AB, Strandbergsgatan 49, S-112 87 Stockholm, Sweden (P.L.).


```

1  TCCGCAATA TGCAGATTA CCGGCCGGT CCGCTCTGAA GCGAGCGCG GGAGGCGCGG
61  CGCGCGCGG CAGCACCAG AACGACCGA GGAAGAAGC CAGCCCCGC CCTCGGCCG
121 TTCCGTCCC ACCCCATCC CGCGGCCCA GGAGGCTCC CCGCTGGCG CGCACTCCCT
181 GTTTCTCTC CTCTGGGCT GGGCTGCTC CCGCTCCGA CTCAGTCTC GCGGGCGCG
241 GTCGCGCAG TCGGTGCTC CCGCGCCAC CTCTCCGGG CCGCGCTCC TAAGGATGG
301 TACTGATTT CGCGGCCCA GGAGACCGG TGGAGCGCG CCGCGCGCG TCGCTCTCC

1
Met Arg Thr Leu Ala Cys Leu Leu Leu
361 TCCGAGCAG CAGCGCTCG GGAGCGG ATG AGG ACC TTG GCT TGC CTG CTG CTC

10 20
Leu Gly Cys Gly Tyr Leu Ala His Val Leu Ala Glu Glu Ala Glu Ile Pro
415 CTC GGC TGC GGA TAC CTC GCC CAT GTT CTG GCC GAG GAA GGC GAG ATC CCG

30 40
Arg Glu Val Ile Glu Arg Leu Ala Arg Ser Gln Ile His Ser Ile Arg Asp
466 CGC GAG GTG ATC GAG AGG CTG GGC CGC AGT CAG ATC CAC AGC ATC CGG GAC

50 60
Leu Gln Arg Leu Leu Glu Ile Asp Ser Val Gly Ser Glu Asp Ser Leu Asp
517 CTC CAG CGA CTC CTG GAG ATA GAC TCC GTA GGG AGT GAG GAT TCT TTG GAC

SstI 70
Thr Ser Leu Arg Ala His Gly Val His Ala Thr Lys His Val Pro Glu Lys
568 ACC AGC CTG AGA GCT CAC GGG GTC CAT GCC ACT AAG CAT GTG CCC GAG AAG

80 90
Arg Pro Leu Pro Ile Arg Arg Lys Arg Ser Ile Glu Glu Ala Val Pro Ala
619 CGG CCC CTG CCC ATT CGG AGG AAG AGA AGC ATC GAG GAA GGT GTC CCC GCT
PDGF-A-1/C* *** **G **C **T ***A

100 110
Val Cys Lys Thr Arg Thr Val Ile Tyr Glu Ile Pro Arg Ser Gln Val Asp
670 GTC TGC AAG ACC AGG ACG GTC ATT TAC GAG ATT CCT CGG AGT CAG GTC GAC
**G *** **C **C **G **C **T *** **C **C **C TCC *** **G ***

120 130
Pro Thr Ser Ala Asn Phe Leu Ile Trp Pro Pro Cys Val Glu Val Lys Arg
721 CCC ACG TCC GCC AAC TTC CTG ATC TGG CCC CCG TGC GTG GAG GTG AAA CCG
*** **C *** **/ PDGF-A-2/**G ***

130 140
Cys Thr Gly Cys Cys Asn Thr Ser Ser Val Lys Cys Gln Pro Ser Arg Val
772 TGC ACC GGC TGC TGC AAC ACG AGC AGT GTC AAG TGC CAG CCG CCG TCC CGC CTG
*** *** **C *** **C *** **C **/

150 160
His His Arg Ser Val Lys Val Ala Lys Val Glu Tyr Val Arg Lys Lys Pro
823 CAC CAC CGC AGC GTC AAG GTG GCC AAG GTG GAA TAC GTC AGG AAG AAG CCA

170 180
Lys Leu Lys Glu Val Gln Val Arg Leu Glu Glu His Leu Glu Cys Ala Cys
874 AAA TTA AAA GAA GTC CAG GTG AGG TTA GAG GAG CAT TTG GAG TGC GCC TGC

180 190
Ala Thr Thr Ser Leu Asn Pro Asp Tyr Arg Glu Glu Asp Thr Gly Arg Pro
925 GCG ACC ACA AGC CTG AAT CCG GAT TAT CGG GAA GAG GAC ACG GGA AGG CCT

200 210
Arg Glu Ser Gly Lys Lys Arg Lys Arg Lys Arg Leu Lys Pro Thr
976 AGG GAG TCA GGT AAA AAA CGG AAA AGA AAA AGG TTA AAA CCC ACC TAA AGC

1027 AGCCAACCG ATGTGAGGTG AGGATGAGCC GCAGCCCTTT CCTGGGACAT GGATGTACAT
1087 GCGGTGTTAC ATTCTGTAAC CTACTATGTA CGGTGCTTTA TTCCAGTGT GCGGTCTTTG
1147 TTCTCTCCG TGAAAACTG TGTCGAGAA CACTCGGGAG AACAAAGAGA CAGTGCACAT
HindIII
1207 TTGTTTAATG TGACATCAAA GCAAGTATT TAGCACTCGG TGAAGCAGTA AGAAGCTTC
1267 TTGTCAAAAA GAGAGAGAGA GAAAGAAAA AAAAGGAAT TC

```

Fig. 1 Nucleotide sequence and deduced amino-acid sequence of the PDGF A-chain determined from a 1.3-kb cDNA clone (D1). An in-frame termination codon in the 5'-untranslated region is underlined. The PDGF A-chain cDNA encodes a 211-amino-acid precursor. Confirmed stretches of PDGF A-chain amino-acid sequence (from ref. 2) are boxed and differences indicated with dashed lines. Restriction endonuclease recognition sites used in the sequencing procedure are indicated. The sequences at which the two oligonucleotide probes PDGF-A-1 and PDGF-A-2 used to identify PDGF A-chain cDNAs were directed are indicated: * implies identity to the cDNA sequence. Box indicates termination codon.

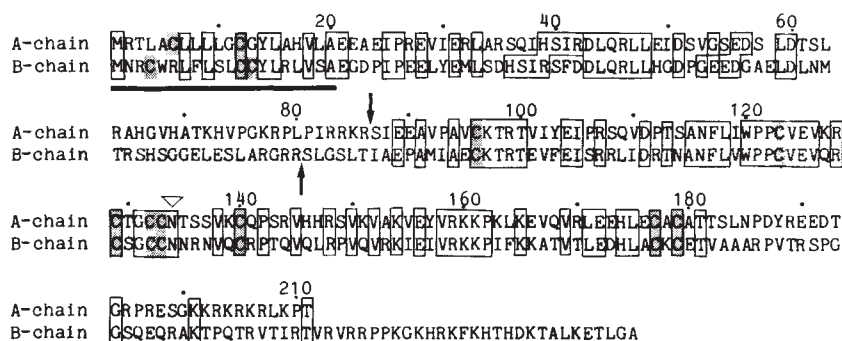
Methods. Standard molecular biology techniques were used where not otherwise indicated. The double-stranded DNA probe PDGF-A-1 was synthesized as two overlapping 50-bp oligonucleotides and radiolabelled using [α - 32 P]-deoxynucleoside triphosphates and the Klenow fragment of DNA polymerase I. PDGF-A-2 was synthesized as a 37-base template and a 12-base complementary primer and was radiolabelled as PDGF-A-1. Both oligonucleotides were synthesized using solid-phase phosphoramidite methodology²¹. The human clonal glioma cell line U-343 MGaC12:6 was the source of poly(A)⁺ RNA, which was prepared using the LiCl/urea method modified as described elsewhere¹¹. Oligo(dT)-primed synthesis of double-stranded cDNA was performed according to Gubler and Hoffman²². The resulting cDNA was treated with T4 DNA polymerase (Biolabs) and subcloned into *Eco*RI-cleaved λ gt10 using *Eco*RI linkers. The recombinant phage were plated on *Escherichia coli* C600 hfl. Duplicate nitrocellulose filter lifts were hybridized with 32 P-labelled oligonucleotide probes at 42 °C in 20% formamide, 5 $\times$ SSC, 50 mM sodium phosphate pH 7.0, 5 $\times$ Denhardt's, 0.1% SDS, 200 μ g ml⁻¹ sonicated salmon sperm DNA and washed in 0.5 $\times$ SSC, 0.1% SDS at the same temperature. The nucleotide sequence of the PDGF A-chain cDNA restriction fragments was determined by dideoxynucleotide chain termination after subcloning into M13 phage derivatives.

The protein sequence matches that derived by amino-acid sequencing of the PDGF A-chain² except at amino acids 119, 141 and 143, found to be Ile, Gln and Ser, respectively, instead of the previously assigned Val, Arg and Thr (Fig. 1). These discrepancies could be due to protein sequencing errors. Alternatively, as the cDNA was obtained from a tumour cell line, it is possible that the sequence deviates from that of the normal PDGF A-chain transcript. The ATG codon at position 388 precedes a basic amino acid (Arg) followed by 18 hydrophobic residues (Fig. 1). This is characteristic of a signal peptide sequence and is consistent with the observation that PDGF A-chain homodimers produced by human osteosarcoma cells are secreted^{25,26}. Comparison with preferred signal peptidase cleavage sites²⁸ suggests that processing may occur between amino acids Ala 20 and Glu 21. The N-terminal sequence of platelet PDGF A-chain is found at amino acid 87, indicating that a propeptide of 66 amino acids (44% charged residues) is cleaved from the precursor to generate a 125-amino-acid A-chain protein. This cleavage occurs after a run of four basic amino acids, Arg-Arg-Lys-Arg. Additional proteolytic processing may occur in the C-terminal region.

Human PDGF is heterogeneous in relative molecular mass, probably reflecting proteolytic cleavage in the platelets or degradation during the purification procedure. SDS-gel electrophoresis of the two constituent chains has revealed that the variability is confined mainly to the A-chain^{1,4}. As amino-acid sequencing showed a unique A-chain N-terminus^{2,4}, this heterogeneity may arise through proteolysis in the highly basic C-terminus (Fig. 1). After N-terminal modification, the A-chain would have a M_r of ~14,000, although the highest- M_r form of the A-chain migrates as a 16–18K species on SDS gels^{1,4}. The discrepancy may be due to glycosylation and/or the anomalous migration commonly observed for cationic proteins. A single consensus sequence for asparagine-linked glycosylation (Asn-X-Ser/Thr) is found at position 134–136, consistent with the report that PDGF contains carbohydrate²⁹. The mature B-chain does not possess any N-glycosylation sites, although one is present in the N-terminal propeptide (Fig. 2).

The 5'-untranslated region of the PDGF A-chain messenger RNA has a high G+C content (~75%) and a high proportion of CpG dinucleotides. CpG-rich regions are found at the 5' end of many vertebrate genes and may indicate that these regions

Fig. 2 Alignment and comparison of the two PDGF chain precursor amino-acid sequences. Homologies are boxed. Cysteine residues are shaded. Signal sequences are underlined and *N*-glycosylation sites are marked with a ∇ . *N*-terminal processing sites are marked with arrows.



are protected from methylation³⁰. Clone D1 carries 281 bp of 3'-untranslated sequence ending with a (dGA)₆ repeat followed by a short poly(A) stretch and *Eco*RI linker but no recognizable polyadenylation signal. Of four cDNA clones sequenced, three terminate around this same position but a fourth contains a longer dGA repeat, extends 370 bp farther 3', but also lacks a polyadenylation signal and poly(A) tail (not shown). It is possible that the three similar clones, including D1, are primed internally on an oligo(A) stretch and represent a mRNA spliced differently from the clone with the longer 3' extension, a possibility in agreement with the presence of multiple A-chain transcripts (Fig. 3). The exact relationship between the different clones and mRNAs remains unknown, although cDNAs in which bases 968–1,036 (Fig. 1) are deleted have been identified (data not shown) and are believed to be the result of differential splicing. If translated, these clones predict an A-chain precursor 15 residues smaller and lacking the basic C-terminal region.

Relationship with the PDGF B-chain

Comparison of the amino-acid sequences of the PDGF A- and B-chain precursors shows them to be similar in size, with an overall amino-acid sequence homology of 40% after insertion of several gaps in their N-terminal portions. A significantly higher degree of homology is seen in a region within the mature chains; amino acids 89–181 of the A-chain is 56% homologous to the B-chain (Fig. 2). Notably, all eight cysteine residues are conserved within the mature chains, implying a similar tertiary structure. Accordingly, homodimers of either the B- or A-chain can bind to the PDGF receptor. The basic region Val-Arg-Lys-Lys-Pro (amino acids 158–162) may be relevant in this context, since basic polypeptides such as protamine sulphate and polylysine have been shown to compete with ¹²⁵I-PDGF for binding to the PDGF receptor³¹.

A significant degree of homology is also seen between part of the N-terminal propeptide sequences, particularly a 10-amino-acid stretch at position 39–48 in the A-chain precursor.

The analogous region in *v-sis* is not essential for this gene's transforming function³². In addition, apparently identical 24K B-chain dimers were formed in NIH 3T3 cells transfected with *v-sis* constructs with or without the N-terminal propeptide region³². Thus, it is difficult to assign a role for this region in post-translational processing of the two PDGF chains.

While there is essentially no sequence homology between the precursor C-terminal sequences, both contain a high proportion of basic amino-acid residues (Fig. 2). Significant nucleotide sequence homology between the A- and B-chain transcripts is observed only in those regions where the amino-acid sequence is strongly conserved.

Hydrophobicity plots (data not shown) indicate that the A- and B-chain precursors are hydrophilic proteins with two major conserved hydrophobic domains. The first corresponds to the signal sequences, while the second is located 28 and 34 residues from the N-terminus of the processed A- and B-chain respectively (Fig. 2) and coincides with a 12-amino-acid conserved region in which there is only one difference (Ile/Val) between the two proteins.

Chromosomal localization

Using 36 human-mouse somatic cell hybrids, we mapped the PDGF A-chain gene to the pter→q22 region of chromosome 7 (Table 1). No other growth factor genes have been localized to this chromosome. The PDGF B-chain gene (*c-sis*) has been mapped⁷ to the long arm of chromosome 22. Interestingly, after duplication of the ancestral PDGF gene, the A- and B-chain genes have acquired different chromosomal localizations.

PDGF mRNA expression in tumour cells

Northern blot hybridization analysis using poly(A)⁺ RNA from various human cell types shows that the PDGF A-chain mRNA is expressed in several of the transformed cell lines examined but is not found in normal human fibroblasts or freshly isolated

Table 1 Distribution of the PDGF A-chain gene with human chromosomes in human-mouse cell hybrids

		Human chromosome																							
	PDGF/Chrom.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
No. of concordant hybrids	(+/+)	9	12	13	12	11	10	19	12	5	16	13	13	12	16	12	8	17	12	9	12	15	8	10	
	(-/-)	19	15	8	13	10	15	18	9	17	9	14	10	12	9	13	15	6	8	18	8	3	12	7	
No. of discordant hybrids	(+/-)	9	7	7	8	9	10	0	8	15	4	6	7	8	4	7	12	3	8	11	8	5	11	6	
	(-/+)	0	4	8	6	9	4	0	10	1	10	5	9	7	10	6	4	12	11	1	11	16	6	10	
% Discordancy		24	29	42	36	46	36	0	46	42	36	29	41	38	36	34	41	39	49	31	49	54	46	48	

A PDGF A-chain cDNA probe (clone D1) was hybridized to Southern blots containing *Eco*RI- or *Hind*III-digested DNA from human-mouse hybrids. Presence of the human PDGF A-chain gene in the hybrids was determined by scoring the presence or absence of human bands on the blots. The first symbol in the parentheses indicates hybrids that were either positive (+) or negative (-) for the PDGF A-chain gene, while the second symbol indicates hybrids that either contained (+) or lacked (-) the particular chromosome. Concordant hybrids have either retained or lost the PDGF A-chain gene together with a specific human chromosome. Discordant hybrids either retained the gene, but not a specific chromosome, or the reverse. Per cent discordancy indicates the degree of discordant segregation of the PDGF A-chain gene and a chromosome. A 0% discordancy is the basis for chromosome assignment. One hybrid, JSR-17S, with a 7/9 translocation, indicates that the PDGF A-chain gene is localized to the pter→q22 region of chromosome 7. The table is compiled from 39 cell hybrids involving 14 unrelated human cell lines and 4 mouse cell lines^{41,42}. The hybrids were characterized by chromosome analysis, by mapped enzyme markers and partly by mapped DNA probes.

Table 2 Compiled data on the expression of the PDGF A- and B-chain genes and secretion of PDGF-like growth factors by human tumour cell lines and normal cells

Cell line	B-chain mRNA	A-chain mRNA	Secretion of a 31K PDGF-like protein	PDGF-receptor competing activity (ng ml ⁻¹)	Mitogenic activity in conditioned medium inhibitable by PDGF antibodies
Tumour cells					
U-2 OS	+	++	++	10	+
U-4 SS	++	++	++	10	+
U-393 OS*	-	+	+	3	ND
SAOS-II	+	-	-	0	ND
SKLMS	-	+	+	2	ND
B-5 GT	-	++	++	15	+
B-6 FS	-	+	+	2	ND
RD	-	++	++	25	+
U-343 MGaC12:6	++	+++	+++	40	+
U-563 MG*	-	-	-	0	ND
Normal cells					
AG 1523	-	-	-	0	-
Macrophages	-	-	ND	ND	ND

The cell lines have the following origins: U-2 OS, osteosarcoma⁴³; U-4 SS, synovial sarcoma⁴³; U-393 OS, osteosarcoma; SAOS-II, osteosarcoma⁴⁴; SKLMS, leiomyosarcoma⁴⁴; B-5 GT, giant cell sarcoma⁴⁵; B-6 FS, fibrosarcoma⁴⁵; RD, rhabdomyosarcoma⁴⁶; U-343 MGaC12:6, glioma⁴⁷; U-563 MG, glioma; AG 1523, a human foreskin fibroblast line obtained from the Human Genetic Mutant Cell Repository, Camden, New Jersey. Macrophage RNA was prepared from freshly isolated peritoneal macrophages, collected by centrifugation of dialysis fluid (1,500g, 5 min). +/− Indicates presence/absence of hybridizing mRNAs on Northern blots or specifically immunoprecipitated 31K proteins that become converted to 16.5–17K species on reduction. PDGF receptor competing activity of serum-free tumour cell-conditioned medium was measured as inhibition of the binding of added ¹²⁵I-labelled PDGF to human foreskin fibroblasts^{25,33,48}. Using a standard curve constructed from results obtained with pure unlabelled PDGF (5–200 ng ml⁻¹), PDGF receptor competing activity of the samples was converted to PDGF equivalents (ng ml⁻¹). Determination of mitogenic activity in serum-free tumour cell-conditioned medium was performed as described previously⁴⁹ in the absence or presence of 50 µg ml⁻¹ of anti-PDGF IgG⁵⁰. ND, not determined.

* Unpublished cell lines of Department of Pathology, Uppsala, Sweden.

peritoneal macrophages (Fig. 3). The macrophages we used were not activated *in vitro* before RNA preparation; after activation, macrophages have been found to express *c-sis* and produce a PDGF-like growth factor^{14,15}. All positive cell lines display three major hybridizing bands, corresponding to transcripts of 1.9, 2.3 and 2.8 kb. Certain human tumour cell lines have been reported to express the PDGF B-chain (*c-sis*) transcript^{8–11}. Some of the cell lines investigated here, such as the glioma U-343 MGaC12:6 and the osteosarcoma line U-2 OS, express both types of transcripts, whereas other cell lines, such as the rhabdomyosarcoma RD and the giant cell sarcoma B-5 GT, express only the A-chain mRNA, and the glioma U-563 MG, like normal fibroblasts and macrophages, expresses none. The A- and B-chain genes are thus regulated independently in human tumour cell lines examined.

Secretion of PDGF-like growth factors

The synthesis of PDGF-like growth factors by human tumour cell lines has been extensively reported^{9–11,20,25,26,33–35}. These factors are all 31K proteins, split by reduction into two closely migrating 16.5K and 17K bands (Fig. 4)^{10,25,33}. They possess the biological features of PDGF and are recognized by anti-PDGF antibodies. Our data show that immunoprecipitation of PDGF-like proteins from the conditioned medium of the human tumour cell lines studied correlates with the expression of PDGF A-chain but not B-chain mRNA (Table 2). This suggests that all of the PDGF-like factors detected by anti-PDGF antibodies in the medium of these human tumour cell lines are composed of only PDGF A-chains, despite the fact that some express both A- and B-chain mRNA (Fig. 3). This view is supported by the detailed structural characterization of the 31K factor secreted by U-2 OS cells which showed it to be an A-chain homodimer²⁶.

Discussion

Our study shows that the two constituent chains of human PDGF are encoded by genes located on different chromosomes, and

that both genes can be expressed independently in human tumour cell lines.

PDGF is stored in the platelet α -granules and released in conjunction with the platelet release reaction (reviewed in ref. 3). It is believed to act as a mitogen for connective tissue cells at the site of vascular injury. Homodimers of both PDGF A- and B-chains possess PDGF receptor agonist activity. What, therefore, is the significance of the presence of both types of chains in human PDGF? Studies of the B-chain homodimer encoded by *v-sis* and the osteosarcoma-derived A-chain homodimer have revealed differences in the efficiency of secretion and/or affinity

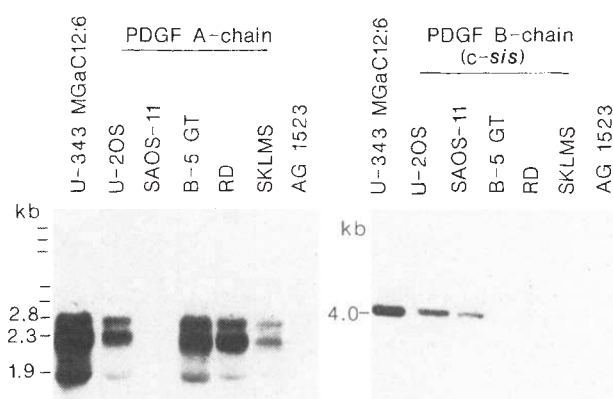
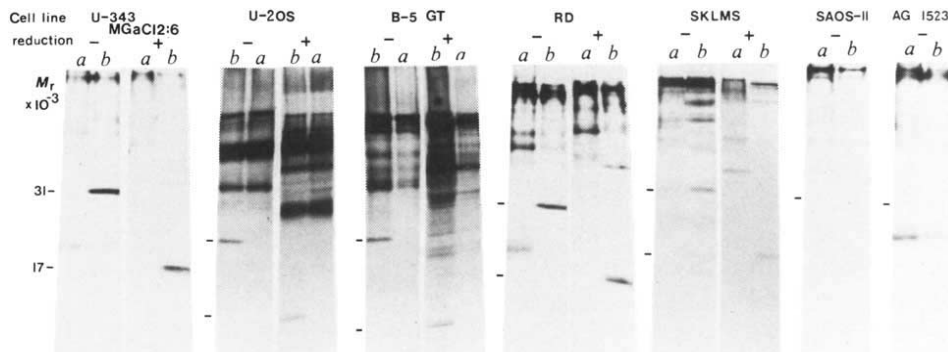


Fig. 3 Northern blot analysis of poly(A)⁺ RNA (10 µg per lane) from various normal and neoplastic human cells. The origins of the cell lines are given in Table 2 legend. Cells growing in monolayer were collected at confluency. Total cellular RNA was prepared and selected once on oligo(dT)-cellulose (Pharmacia). Agarose gel electrophoresis, blotting to nitrocellulose and hybridization to ³²P-labelled PDGF A-chain cDNA (left) or PDGF B-chain (*c-sis*) cDNA (right) were performed as described previously¹¹. Filters were exposed to Kodak XAR-5 films at −70 °C for 4 days.

Fig. 4 Immunoprecipitation of metabolically labelled PDGF-like growth factors produced by human tumour cells. The origins of the cell lines are given Table 2 legend. Confluent 850-cm² roller bottle cultures of cells were labelled with ³⁵S-cysteine (NEN, 600 Ci mmol⁻¹) as described elsewhere²⁵. Briefly, cultures were pulsed with 250 µCi of ³⁵S-cysteine in 3 ml of cysteine-free medium for 3 h, and then chased in 3 ml of cysteine-containing medium for an additional 3 h. Media were pooled and sequentially precipitated with a control rabbit serum (a) and PDGF antiserum (b). Immunoprecipitates were absorbed to protein A-Sepharose (Pharmacia) and analysed on 13–18% SDS-polyacrylamide gels under reducing or non-reducing conditions. Dried gels were exposed to Kodak XAR-5 film for 4–7 days at –70 °C.



for a specific cellular compartment^{25,36}, raising the possibility that structural differences between the two chains serve different functions in relation to storage, release and association with the plasma membrane, extracellular matrix and plasma proteins^{37–39}. For example, both types of homodimer are biologically active but their affinity for the PDGF receptor may differ both from each other and from the putative heterodimer. In fact, platelet PDGF appears to be more potent than the PDGF-like factors purified from human tumour cell line-conditioned media (C.-H.H. *et al.*, unpublished results). Furthermore, in spite of evidence that the transforming function of SSV is exerted by an externalized *v-sis* product, no accumulation of PDGF agonist activity is seen in the medium of acutely SSV-transformed human fibroblast cultures, and anti-PDGF antibodies precipitate only low-*M_r* monomers from the medium of SSV-transformed cells²³. Apparently, after being released, the *v-sis* product remains associated with, or rapidly associates with, structures in the cell membrane including the PDGF receptor^{18–36}. The low-*M_r* monomers probably represent degraded *v-sis* products. In contrast, intact 31K A-chain homodimers can be immunoprecipitated from human tumour cell-conditioned medium. The A-chain may therefore contribute to the stability of PDGF.

The exact nature of the human PDGF subunit composition and the significance of the presence of both A- and B-chains remain unknown. The genetic basis for A-chain expression in human tumour cells is also unknown, as is its role in tumour growth. Several non-transformed cell types, endothelial cells¹², cytotrophoblasts¹³, smooth muscle cells⁴⁰ and activated macrophages^{14,15} have been shown to express the *c-sis* gene and/or to release PDGF-like growth factors. It will be interesting to see whether the A-chain gene is expressed in these normal cells, and to determine the subunit structure of the secreted factors. Knowledge of the PDGF A-chain precursor structure and access to PDGF A-chain cDNAs as molecular probes will certainly contribute to the elucidation of such matters.

We thank Monica Nistér for providing the cell line U-343 MGaC12:6, Peter Wenkler for help with oligonucleotide synthesis, Flossie Wong-Staal for the gift of a *c-sis* cDNA clone, and Rolf Ohlsson and Margaret Bywater for valuable contributions during the initial phase of this work. The studies were supported in part by grants from the Swedish Cancer Society. Konung Gustaf V:s åttioårsfond, the Swedish Department of Agriculture (B.W. and C.-H.H.), NIH grants GM 20454 and HD 05196 and American Cancer Society grant CD62 (T.B.S.).

Received 24 January; accepted 13 March 1986.

1. Johnsson, A., Heldin, C.-H., Westermark, B. & Wasteson, Å. *Biochem. biophys. Res. Commun.* **104**, 66–74 (1982).
2. Johnsson, A. *et al.* *EMBO J.* **3**, 921–928 (1984).
3. Heldin, C.-H., Wasteson, Å. & Westermark, B. *Molec. cell. Endocr.* **39**, 169–187 (1985).
4. Waterfield, M. D. *et al.* *Nature* **304**, 35–39 (1983).
5. Doolittle, R. F. *et al.* *Science* **221**, 275–277 (1983).
6. Josephs, S. F., Guo, C., Ratner, L. & Wong-Staal, F. *Science* **223**, 487–490 (1984).
7. Swan, D. C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 4691–4695 (1982).
8. Eva, A. *et al.* *Nature* **295**, 116–119 (1982).
9. Pantazis, P., Pellicci, P. G., Dalla-Favera, R. & Antoniades, H. N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2404–2408 (1985).
10. Graves, D. T. *et al.* *Science* **226**, 972–974 (1984).
11. Betsholtz, C., Westermark, B., Ek, B. & Heldin, C.-H. *Cell* **39**, 447–457 (1984).
12. Collins, T., Ginsburg, D., Boss, J. M., Orkin, S. H. & Pober, J. S. *Nature* **316**, 748–750 (1985).
13. Goustin, A. S. *et al.* *Cell* **41**, 301–312 (1985).
14. Shimokado, K. *et al.* *Cell* **43**, 277–286 (1985).
15. Martinet, Y. *et al.* *Nature* **319**, 158–160 (1986).
16. Robbins, K. C., Antoniades, H. N., Devare, S. G., Hunkapiller, M. W. & Aaronson, S. A. *Nature* **305**, 605–608 (1983).
17. Deuel, T. F., Huang, J. S., Huang, S. S., Stroobant, P. & Waterfield, M. *Science* **221**, 1348–1350 (1983).
18. Huang, J. S., Huang, S. S. & Deuel, T. F. *Cell* **39**, 79–87 (1984).
19. Owen, A. J., Pantazis, P. & Antoniades, H. N. *Science* **225**, 54–56 (1984).
20. Bowen-Pope, D. F., Vogel, A. & Ross, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2396–2400 (1984).
21. Garrett, J. S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 7466–7470 (1984).
22. Johnsson, A., Betsholtz, C., von der Helm, K., Heldin, C.-H. & Westermark, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1721–1725 (1985).
23. Johnsson, A., Betsholtz, C., Heldin, C.-H. & Westermark, B. *Nature* **317**, 438–440 (1985).
24. Stroobant, P. & Waterfield, M. D. *EMBO J.* **3**, 2963–2967 (1984).
25. Betsholtz, C. *et al.* *Biochem. biophys. Res. Commun.* **117**, 176–182 (1983).
26. Heldin, C.-H. *et al.* *Nature* **319**, 511–514 (1986).
27. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).

28. Von Heijne, G. *J. molec. Biol.* **173**, 243–251 (1984).
29. Deuel, T. F. *et al.* *J. biol. Chem.* **256**, 8896–8899 (1981).
30. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. *Cell* **40**, 91–99 (1985).
31. Huang, J. S., Huang, S. S., Kennedy, B. & Deuel, T. F. *J. biol. Chem.* **257**, 8130–8136 (1982).
32. King, C. R., Giese, N. A., Robbins, K. C. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5295–5299 (1985).
33. Nistér, M., Heldin, C.-H., Wasteson, Å. & Westermark, B. *Proc. natn. Acad. Sci. U.S.A.* **81**, 926–930 (1984).
34. Gudas, L. J., Singh, J. P. & Stiles, C. D. *Cold Spring Harb. Conf. Cell Proliferation* **10**, 229–236 (1983).
35. Rozengurt, E., Sinnett-Smith, J. & Taylor-Papadimitriou, J. *Int. J. Cancer* **36**, 247–252 (1985).
36. Robbins, K. C., Leal, F., Pierce, J. H. & Aaronson, S. A. *EMBO J.* **4**, 1783–1792 (1985).
37. Smith, J. C., Singh, J. P., Lillquist, J. S., Goon, D. S. & Stiles, C. D. *Nature* **296**, 154–156 (1982).
38. Raines, E. W., Bowen-Pope, D. F. & Ross, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3664–3668 (1981).
39. Huang, J. S., Huang, S. S. & Deuel, T. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 343–346 (1984).
40. Nisson, J., Sjöberg, M., Palmberg, L., Thyberg, J. & Heldin, C.-H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4418–4422 (1985).
41. Sakaguchi, A. Y. & Shows, T. B. *Somat. Cell Genet.* **8**, 83–94 (1982).
42. Shows, T. B. *et al.* *Somat. Cell molec. Genet.* **10**, 315–318 (1984).
43. Pontén, J. & Saksela, E. *Int. J. Cancer* **2**, 434–447 (1967).
44. Fogh, J. & Trempe, G. in *Human Tumor Cells in Vitro* (ed. Fogh, J.) 115–159 (Plenum, New York, 1975).
45. Thurzo, V. *et al.* *Neoplasma* **23**, 577–587 (1976).
46. McAllister, R. M., Melnyk, J., Finkelstein, J. Z., Adams, E. C. & Gardner, M. B. *Cancer* **24**, 520–526 (1969).
47. Pontén, J. & Westermark, B. *Med. Biol.* **56**, 184–193 (1978).
48. Heldin, C.-H., Westermark, B. & Wasteson, Å. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3664–3668 (1981).
49. Betsholtz, C. & Westermark, B. *J. cell. Physiol.* **118**, 203–210 (1984).
50. Heldin, C.-H., Westermark, B. & Wasteson, Å. *Expl. Cell Res.* **136**, 255–261 (1981).
51. Warner, B. D. *et al.* *DNA* **3**, 401–411 (1984).
52. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263–269 (1983).

The AMPTE artificial comet experiments

A. Valenzuela, G. Haerendel, H. Föppl, F. Melzner, H. Neuss,
E. Rieger, J. Stöcker, O. Bauer, H. Höfner & J. Loidl

Max-Planck-Institut für Physik und Astrophysik, Institut für extraterrestrische Physik, 8046 Garching, FRG

In July last year and in December 1984, barium clouds were injected into the solar wind from the AMPTE satellite. The clouds resembled those of natural comets in that a head and tail were clearly visible. Tail rays formed with speeds of a few tens of kilometres per second, and other structures separated at even higher speeds. The comet head, instead of moving slowly downstream, performed a lateral excursion of several hundred kilometres. In the first of a series of six articles, the properties of the 'artificial comet' are described.

At the time of writing, four space probes are on their way to encounter comet Halley in March 1986. Another spacecraft recently intercepted the tail of comet Giacobini-Zinner. At about the same time, a plan conceived by L. Biermann and colleagues in 1961¹, to simulate the plasma aspects of solar wind-comet interactions, has been achieved. By means of barium vapour releases in the solar wind, several phenomena were created which are expected to occur in comets, some of which strikingly resemble the visual appearance of a comet. Thus the name 'artificial comet', chosen for these experiments, appears to be appropriate, even though only the plasma interactions can be studied. Another difference is that the spatial scales of the interaction region are comparable to the microscale of the physical processes, the ion gyroradius.

Two barium experiments were performed, one at the morning flank of the magnetosphere on 27 December 1984, and the other on the evening flank on 18 July 1985. They constitute the culmination of twenty years of application of this technique by the Max-Planck Institute for Extraterrestrial Physics (MPE), starting with the first barium cloud created in the ionosphere over the Sahara desert in November 1964².

The opportunity for these experiments arose with the implementation of the AMPTE (Active Magnetospheric Particle Tracer Explorers) mission, a cooperative venture between the United States, West Germany and the United Kingdom³. Although the prime aim of the mission was the long-range tracing of the transport of ions from the solar wind into the magnetosphere and further redistribution therein, it was ideally suited for the study of the initial interaction of the injected plasma with the ambient medium. To this end a set of plasma and field diagnostic instruments was installed on the German Ion Release Module (IRM), which carried the barium and lithium release containers⁴. Furthermore, a subsatellite developed in the United Kingdom (UKS) was injected into nearly the same orbit as the IRM in order to provide another probe for the plasma and field perturbations created during the release experiments⁵. In addition to the *in situ* diagnostics, optical observations from the earth contribute valuable information about the dynamics of the solar wind interaction with the seeded plasma.

The set of six articles presented in this issue of *Nature* addresses various aspects of the interaction of the solar wind with the barium plasma in the release of 27 December 1984. The optical phenomena are described in the present paper and that immediately following⁶. The third paper⁷ presents the magnetic field measurements from both the IRM and UKS spacecraft, and the two which follow^{8,9} the plasma measurements on UKS and the wave signatures of the barium plasma and its interaction with the solar wind as observed in IRM and UKS. A final paper¹⁰ attempts to describe the main elements of the interaction dynamics, the *in situ* plasma measurements on the IRM play a decisive diagnostic role.

The main aim of this paper is twofold: to present the experiment characteristics and to document the cometary nature of

the experiments on the basis of optical observations obtained from an aircraft above the southern Pacific Ocean. The physical insight derived from this data set bears on the momentum transfer between two very different, extremely rarified plasmas on length scales comparable to or much smaller than the gyroradii of the interacting ions (H^+ , Ba^+). The greatest surprise was the gross deviation from a simple picture derived from the consideration of two interacting fluids.

Experiment configuration

The barium vapour release on 27 December 1984 occurred outside the Earth's magnetopause at a distance of $17.2 R_E$ (Earth radii), low geographic latitude, at 07:06 local time, 12:32:01 universal time (UT). Photoionization by the sun converted the neutral vapour into a plasma on a timescale of ~ 30 s. Contrary to our planning, the release location was in front of the bow shock; this was caused by an increase of solar wind pressure approximately $1/2$ h before the experiment⁷. The timing of the experiment was dictated by the requirement of simultaneous visibility of the point of injection from the observation sites, installed in favourable places in the Northern Hemisphere (Boulder, Colorado; Kitt Peak National Observatory, Arizona; White Sands and Los Alamos, New Mexico; Mauna Kea and Mt Haleakala, Hawaii and Nau'i¹¹). In addition, two aircraft equipped with highly sensitive cameras were flown over the northern and southern pacific. It turned out that ground visibility was unusually poor on the day, as it had been on the prime release day (25 December, 1984). Only the observers near Boulder had clear sky. Therefore, the success of the experiment rested largely on the aircraft observations.

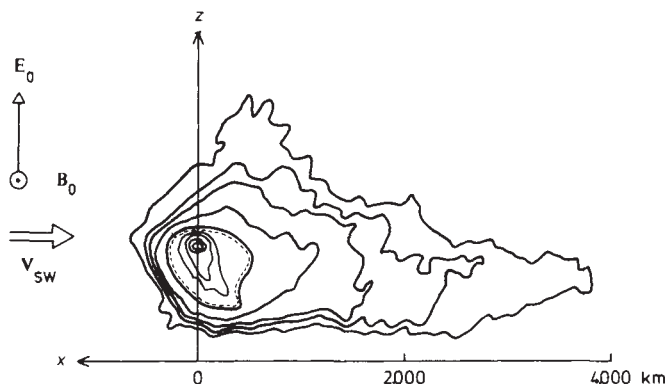


Fig. 1 Combination of brightness contours of images taken at +3.2 min from the South Pacific and with a less sensitive camera from Boulder, Colorado⁶. The comet head (seen from Boulder, and enclosed by the dashed contour) has a density distribution elongated at a large angle to the tail. It fits entirely into the apparent comet head on the pictures taken from the south, which are overexposed and enlarged by blooming. The directions of the ambient electric and magnetic fields, and of the solar wind velocity vector, are indicated.

The data presented in this paper were obtained with a low-light-level television camera installed and run by a team from the MPE aboard an Argentinean Air Force Boeing 707 which was cruising $\sim 2,000$ km south of Tahiti¹¹. The observer's line-of-sight was generally at only a small angle to the undisturbed ambient magnetic field (a few degrees, but varying), and nearly perpendicular to the solar wind flow vector. The plane of projection (see Fig. 1) can therefore be considered to contain the flow vector and the ambient electric field, $\mathbf{E}_0 = -(\mathbf{V}_{sw} \times \mathbf{B}_0)/c$ (where $\mathbf{V}_{sw}$ is the solar wind velocity, $\mathbf{B}_0$ is the ambient magnetic field and c is the speed of light).

The UKS was located at a distance of 170 km from the IRM at an azimuth angle of 120° , measured from the sunward direction of the morning side ($-y$ direction)⁷. The interplanetary plasma and magnetic field parameters ($v_{sw} = 540 \text{ km s}^{-1}$, $B_0 = 10 \text{ nT}$, ion density $n_0 = 5 \text{ cm}^{-3}$) were such that the initially created magnetic cavity did not extend to the UKS. This spacecraft was ideally located to observe the draping of the magnetic field around the comet head.

The second artificial comet (on 18 July 1985) was created at the evening flank of the magnetosphere, this time behind the Earth's bow shock as intended. Hence, the ambient flow velocity was lower ($\sim 370 \text{ km s}^{-1}$) and the density higher ($\sim 15 \text{ cm}^{-3}$) than during the first experiment. This experiment will be discussed elsewhere. Here, we merely show a picture in order to contrast it with the quite different appearance of the first artificial comet.

Optical observations

The observations of the MPE team from the Boeing 707 were hampered by a sudden onset of turbulence at the time of barium injection and imperfect functioning of the computer control of the heliostatic platform used for stabilizing the cloud's image. As a result, apart from a few sporadic intervals of coverage during the first 3.7 min, continuous observation started only at 12:35:43 UT and continued until the barium plasma cloud vanished at 12:37:10 UT. As a result, when continuous observation with this camera began, a long tail had already formed. During the following 80–90 s, dramatic transformations of the tail took place, followed by the final disintegration of head and tail. The exposure times for the television camera (0.32 s) were chosen so as to obtain optimum coverage of the tail, the brightness of which was about two orders of magnitude lower than that of the head. This caused overexposure of the head region and blooming of its image.

Figure 3 shows a picture of the 27 December 1984 artificial comet after 4.4 min. The shape of the core region as it appears on the television images (white region) is severely enlarged and distorted during most of the observation time, until at 12:36:40 UT the brightness was reduced below the saturation level of the television camera. The development of the comet head was therefore much better covered by the cameras on the ground and aboard the aircraft (a NASA Convair 990) in the Northern Hemisphere¹⁶. In order to relate the two sets of observations we have superposed on some smoothed brightness contours observed from the South Pacific the much better resolved contours obtained with the camera of M. Mendillo and J. Baumgardner at Boulder (Fig. 1). Because of lower sensitivity this instrument covered only the bright core or head region of the comet, that is, the region saturated and distorted by blooming on the MPE camera. An interesting feature is the orientation of the contours in the head. They are pointing at about 30° tailward of the $-z$ axis, rather than in the antisolar direction as the tail did when viewed from the South Pacific. In addition, the more sensitive cameras on the NASA Convair⁶ yielded a deviation of the projected long tail from the $-x$ axis by $\sim 25^\circ$ in the direction of $-z$. Precise triangulation of the tail has not yet been possible because of a lack of calibration of the brightness contours. But it appears that there is a substantial flaring of the tail in the x - y plane, that is, approximately in the plane of observation from

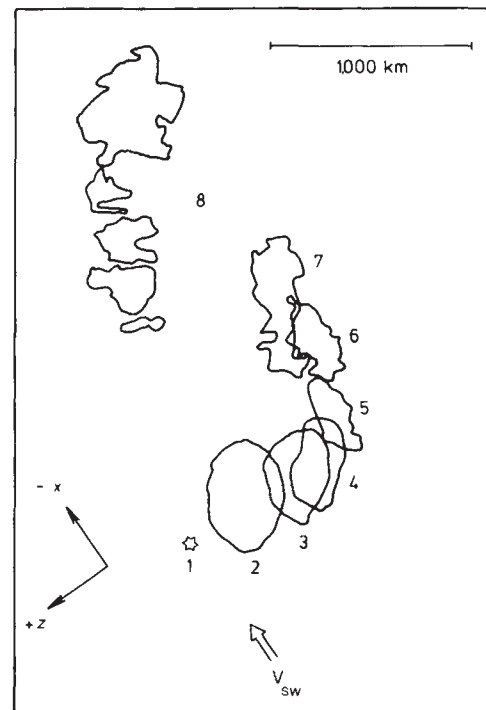


Fig. 2 Outlines of the comet head (initially enlarged by blooming) during the life of the artificial comet of 27 December 1984. Rather than moving in the flow direction of the solar wind, the comet head is displaced sideways during the initial 4.5 min. Outline 1, +4 s; 2, +181 s; 3, +242 s; 4, +256 s; 5, +280 s; 6, 288 s; 7, +294 s; 8, +306 s.

the South Pacific and therefore difficult to assess with any accuracy.

As expected^{12,13}, a magnetic cavity formed in the initial expansion phase of the Ba plasma⁷. Its maximum extent was apparently reached after ~ 60 s (refs 1, 10). This is consistent with the expansion of $\sim 10^{25}$ Ba ions at a mean velocity of 1.35 km s^{-1} if one takes into account the observed compression of the magnetic field. The kinetic pressure of a mass M_i of Ba ions expanding with velocity $\langle v_{exp} \rangle$ is balanced by the magnetic pressure of the compressed ambient field, B_c (high alfvénic Mach number):

$$\frac{M_i}{(4\pi/3)R_{dia}^3} \langle v_{exp} \rangle^2 \approx \frac{B_c^2}{8\pi}$$

Here $R_{dia} = \langle v_{exp} \rangle \tau_{dia}$ is the radius of the cavity and τ_{dia} is the duration of the expansion. With $\tau_{dia} = 60$ s and $M_i = 2 \text{ kg}$ one obtains for the compressed field $B_c = 64 \text{ nT}$. This number is not unreasonable in view of the later observed compression of up to 120 nT (ref. 7). This means that R_{dia} did not exceed ~ 100 km. Rees *et al.*⁶ deduced a radius of 90 km from their optical data. Thus, on the contours shown in Fig. 1, the diamagnetic cavity would occupy only the innermost area of the comet head, had it existed at the time when the picture was taken. But it is strongly suggested that at this time the whole comet head was already pervaded by the magnetic field.

The return of the field was observed to occur at the IRM spacecraft at +77 s (ref. 7) (time after Ba release). The process by which this fast return and compression of $\mathbf{B}$ is achieved will be discussed in ref. 10. In this context it is important to note that substantial momentum transfer from the solar wind to the Ba plasma must have started ~ 1 min after release. The fast development of a visible tail as registered by the northern observers⁶ is consistent with this conclusion.

The greatest surprise with regard to momentum transfer was the observation that the comet head did not move at all in the direction of solar wind flow during the first 4.6 min. Instead, a substantial displacement in the $-z$ direction was observed; that

Fig. 3 Image of the artificial comet of 27 December 1984, 4.4 min after barium plasma injection. Image taken with a low-light level television system installed on an Argentine Boeing 707 flying over the southern Pacific Ocean. The tail is oriented approximately along the flow direction of the solar wind. The comet head is overexposed (white area).

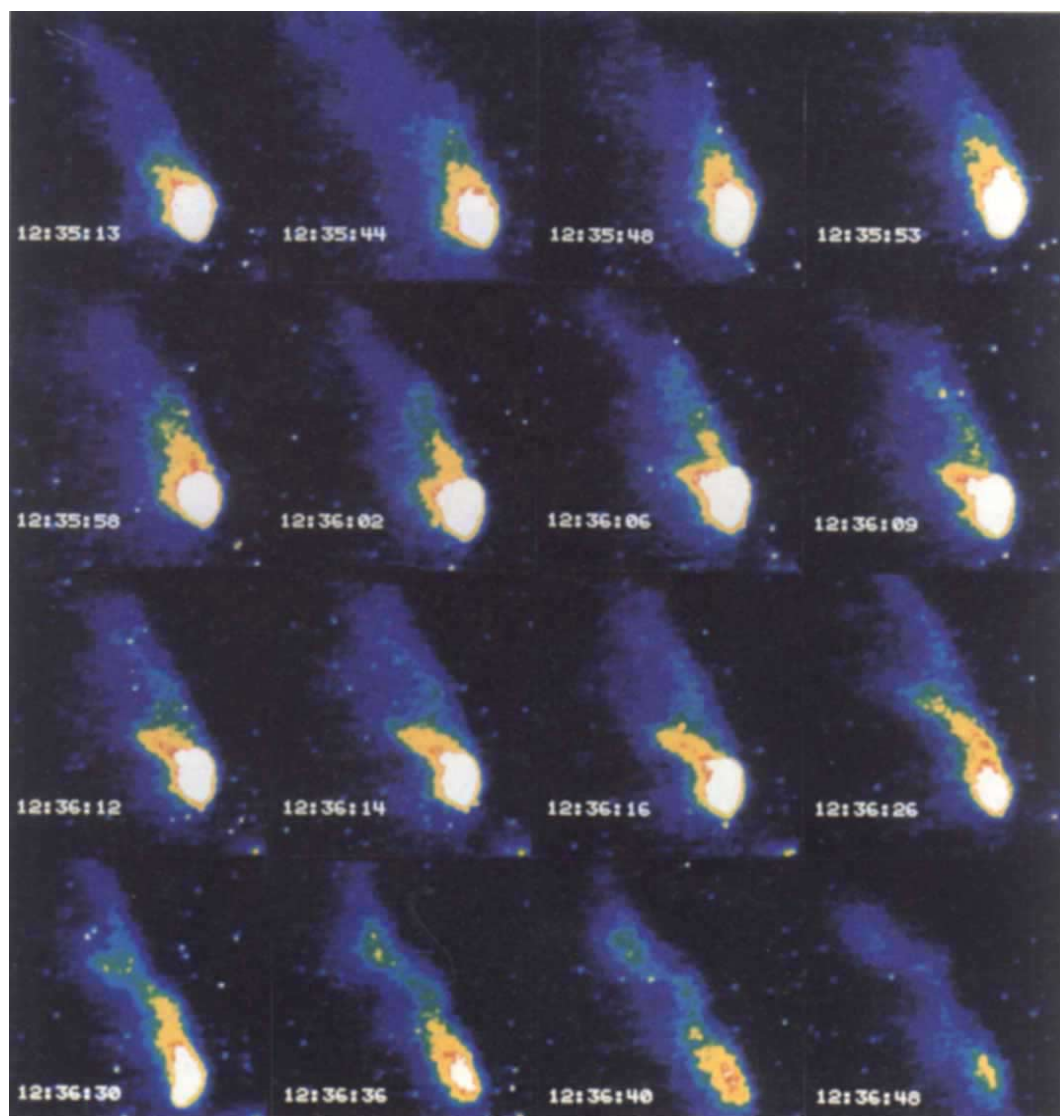
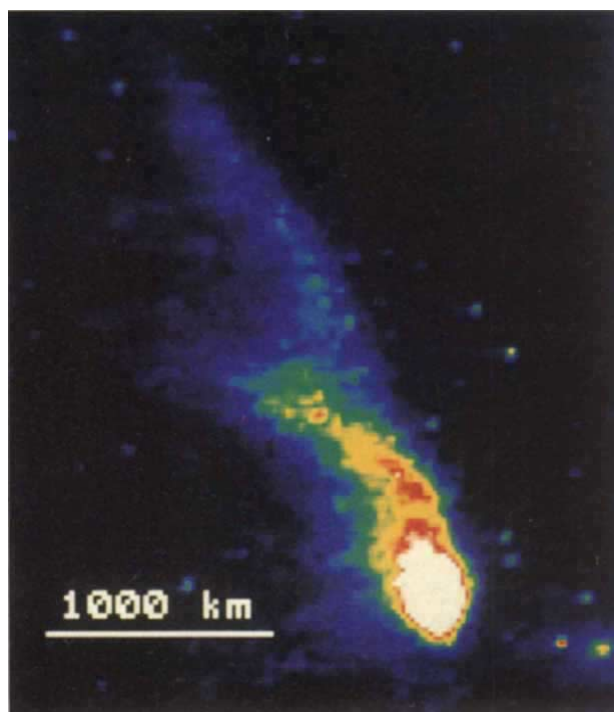


Fig. 4 Sequence of images showing the development of the tail during the last minute of the comet's life. A second tail or tail ray forms and moves downstream.

is, at right angles to the solar wind flow. Only when its density (and brightness) had decayed substantially, did the comet head exhibit the expected acceleration in the solar wind direction. Figure 2 shows the displacements of the core region in x and z with respect to the IRM spacecraft. The trajectory is consistent with a nearly constant acceleration of 20 m s^{-2} during the first 4 min. The displacement in $-z$ exceeded 1,000 km after 5 min.

This completely unexpected displacement transverse to the solar wind velocity vector and to the ambient magnetic field is interpreted¹⁰ as an expression of the recoil of head and tail caused by the extraction of ions in the $+z$ direction by the interplanetary electric field. Signatures of this extraction are the protrusions or 'fingers' of the density distribution extending in the $+z$ direction from head and tail, as well as the apparent motions therein (Fig. 3). The brightness distribution inside the comet head, on the other hand, is a consequence of the recoil. An electric field displaces the ions approximately in the $-z$ direction and, at the same time, drives a current at right angles ($-x$ direction). The corresponding magnetic field gradient manifests itself through a density/brightness modulation.

The absence of motion in the solar wind direction is also a recoil effect. The force (Maxwell stresses) exerted by the solar wind is communicated exclusively to the rear end of the comet head. An electric field applies the force to the ions, which are accelerated so as to form the observed tail¹⁰.

During the period covered by the MPE television system the tail underwent a dramatic development. A second tail or new tail ray was formed. It appeared at 12:35:58 UT (Fig. 4) as a slight deformation on the upper rear side of the comet head. Within the next 15 s the deformation developed into a jet growing with a projected speed of $\sim 35 \text{ km s}^{-1}$ at an angle of $\sim 45^\circ$ between the $+z$ and $-x$ axes. It is interesting to note that during the 20 s preceding its appearance the magnetic field had gone through a rapid change from the low elevation angles (in solar-ecliptic coordinates) that prevailed before and afterwards, through -45° to $+45^\circ$ and back to 0° . It may well be that it was the magnetic flux captured by the head plasma within this period that gave rise to the new tail feature.

Already while this ray was appearing, its base at the comet head was seen to move downstream. By 12:36:26 UT it had clearly separated from the head and continued to move along the direction of the still visible former tail; that is, nearly in the $-x$ direction. The speed of separation reached $\sim 60 \text{ km s}^{-1}$ after 30 s of development and remained nearly constant for the next 30 s. This new tail ray visibly drained the comet head of Ba ions. In Fig. 4 the overexposed high-density core is seen to shrink in size and comes out of saturation at 12:36:40 UT. At this time, the head began the highly accelerated motion in the anti-sunward direction shown in Fig. 2 (outline 5). Head and second tail ray exhibit the same morphological development: a substantial smearing of the density distribution and corresponding decrease in brightness is accompanied by an acceleration in the $+z$ direction, which is the orientation of the unperturbed electric field. Thus, test particle behaviour follows the initial phase characterized by collective interactions. Soon afterwards all visible phenomena had disappeared.

Figure 5 shows the artificial comet of 18 July 1985. The image was taken with a ground-based telescope ($f = 940 \text{ mm}$) and low-light-level television camera¹¹ in Arequipa (Peru). At this stage of development, $\sim 200 \text{ s}$ after plasma injection, the head is composed of several arches, very reminiscent of the structure of natural comet heads (as in drawings based on visual observations). In this experiment the brightness contrast between head and tail was much smaller than on 27 December 1984. The reason is quite probably the much different flow environment behind the bow shock. The alfvénic Mach number of the flow component perpendicular to $\mathbf{B}_0$ was only 2.3, as compared with 6 on 27 December 1984. The ambient density, on the other hand, was three times higher than in the first experiment (15, instead of 5 cm^{-3}). In the accompanying paper¹⁰ we argue that the tail

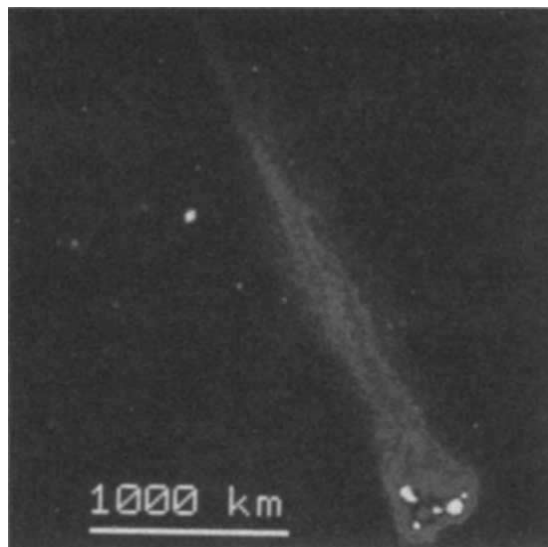


Fig. 5 The artificial comet of 18 July 1985 at +200 s, observed from Arequipa (Peru). The head is composed of several arches.

density should equal that of the ambient plasma; hence the relatively bright tail in the second experiment. Additionally, we later saw the separation of one of the arches that formed in the head. Moving in the antisolar direction at $\sim 100 \text{ km s}^{-1}$, it had the appearance of a disconnection event such as is occasionally observed in the plasma tails of natural comets.

Conclusion

The interaction between the solar wind and an artificially injected plasma cloud exhibits many comet-like features, although the interaction occurs on spatial scales of the order of the thermal ion gyroradius. Head and tail are clearly discernible. The most striking observations are the absence of a noticeable downstream motion of the comet head and an accelerated transverse displacement. This is attributed to a collective reaction or recoil of the head to the ion injection into the tail and the simultaneous extraction of ions from head and tail by the ambient electric field¹⁰. The collective behaviour is enforced on the unmagnetized ions by electric fields, which act to maintain charge neutrality with the magnetized electrons, and also drive the electron Hall currents which to a large extent create the distortion and compression of the magnetic field captured by the cloud¹⁰.

We thank the many colleagues at MPE who contributed to the success of the artificial comet experiments; the German Satellite Operations Center of the DFVLR for support in mission analysis, management and operations; the Bundesministerium für Forschung und Technologie for financial support under grant WRK 01 OM 010-ZA/WF/WRK 275; the Argentine Air Force for providing a Boeing 707 for optical observations, and many observatories in North and South America for their hospitality and technical support.

Received 2 December 1985; accepted 7 March 1986.

1. Biermann, L., Lüst, R., Lüst, R. & Schmidt, H. U. *Z. Astrophys.* **53**, 226–236 (1961).
2. Haerendel, G., Lüst, R. & Rieger, E. *Planet. Space Sci.* **15**, 1–18 (1967).
3. Krimigis, S. M., Haerendel, G., McEntire, R. W., Paschmann, G. & Bryant, D. A. *Eos* **63**, 843–850 (1982).
4. Häusler, B. et al. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 192–201 (1985).
5. Ward, A. K. et al. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 202–211 (1985).
6. Rees, D., Hallinan, T. J., Stenback-Nielsen, H. C., Mendillo, M. & Baumgardner, J. *Nature* **320**, 704–708 (1986).
7. Lüth, H. et al. *Nature* **320**, 708–711 (1986).
8. Rodgers, D. J. et al. *Nature* **320**, 712–716 (1986).
9. Gurnett, D. A. et al. *Geophys. Res. Lett.* **12**, 851–854 (1985).
10. Haerendel, G., Paschmann, G., Baumjohann, W. & Carlson, C. W. *Nature* **320**, 720–723 (1986).
11. Haerendel, G. et al. *IEEE Trans. Geosci. Remote Sensing* **GF-23**, 253–258 (1985).
12. Haerendel, G. *Eur. Space Ag. J.* **7**, 135–144 (1983).
13. Haerendel, G. *Eur. Space Ag. Spec. Publ.* No. SP-195, 337–340 (1983).

Optical observations of the AMPTE artificial comet from the Northern Hemisphere

David Rees*, Thomas J. Hallinan†, Hans C. Stenbaek-Nielsen†, Michael Mendillo‡
& Jeffrey Baumgardner‡

* Department of Physics and Astronomy, University College London, Gower Street, London WC1 6BT, UK

† Geophysical Institute, University of Alaska, Fairbanks, Alaska 99701, USA

‡ Department of Astronomy, Boston University, Boston, Massachusetts 02215, USA

The AMPTE artificial comet of 27 December 1984 developed a well-defined core (maximum size $\sim 160 \times 100$ km) of cold barium ions, surrounded by a diffuse coma, 500 km in diameter, consisting of ions extracted from the core. A plasma tail extended from the core, mainly directed away from the Sun. The plasma tail and diffuse coma consisted of energetic ions (>200 eV). Confounding predictions, the artificial comet survived for barely 5 minutes, and the core of the comet moved southward, rather than in the direction of the solar wind.

AS part of the international AMPTE (Active Magnetospheric Particle Tracer Explorers) mission^{1,2}, several ground-based, widely-spaced observing facilities and two instrumented airborne observatories were deployed for the AMPTE artificial comet experiment on 27 December 1984. Unfortunately, poor weather precluded observations from all but one of the ground-based stations. The data to be described in this report were obtained with the University of Alaska's low-light-level television (LLTV) system and a UCL Doppler imaging system (DIS³), aboard a NASA Convair 990 aircraft (the Galileo II Airborne Observatory, near 38.137° N, 120.268° W at the time of the barium release). In addition, Boston University's Mobile Ionospheric Observatory (MIO) provided ground-based observations using an image-intensified camera system located in Boulder Canyon, Colorado (40.016° N, 105.3344° W). A companion article⁴ describes observations made simultaneously from an Argentinian Boeing 707 instrumented aircraft flying out of Tahiti.

Observations

The initial phase. The gas of neutral barium atoms released into the solar wind was subject to radial expansion, photoionization by solar ultraviolet radiation (with a time constant of 20–30 s), and bulk motion imparted by the Ion Release Module (IRM) orbital speed, as modified (perhaps) by the barium injection mechanism. During this initial phase, unfiltered ('white-light') observations from the LLLTV and the ground-based imaging systems recorded the expansion of the cloud, as revealed by the ensemble of optical lines emitted by the neutral and partially ionized barium, as well as by residual neutrals (mainly strontium) in the gaseous mixture. As shown in Fig. 1, the expanding spherical neutral cloud was essentially coincident with the plasma cloud for ~ 2 min after barium release. As shown in Fig. 2a, the white-light image after 61 s has a radius of 90 km, corresponding to a mean radial expansion speed of ~ 1.5 km s^{-1} . The airborne LLLTV camera revealed a faint spherical cloud (attributed to neutral strontium) expanding at a similar rate (~ 1.2 km s^{-1}) throughout the several minutes of observations. Figure 3 shows examples of the spherical neutral cloud, separated from plasma structures, ~ 3 min after the release.

During the first 30 s of expansion, an ion cloud formed, with an intense central core. Photoionization of the remaining neutral barium continued for the next 30–60 s. The expansion of this plasma, the establishment of dynamical equilibrium with the solar wind, and the formation of a diamagnetic cavity excluding the Interplanetary Magnetic Field (IMF) were measured *in situ* by the IRM and the United Kingdom Satellite (UKS)⁵, the latter being only 170 km from the IRM. The imaging results show that at ~ 90 s after the release ($+90$ s), when photoionization was

essentially complete, the intense core of the ion cloud reached a maximum size of 160×100 km, similar in size to that predicted for the diamagnetic cavity². The size of the ion core remained relatively constant (see Fig. 2b at $+187$ s), with a steadily decreasing total brightness due to the rapid, continuous loss of ions into the solar wind.

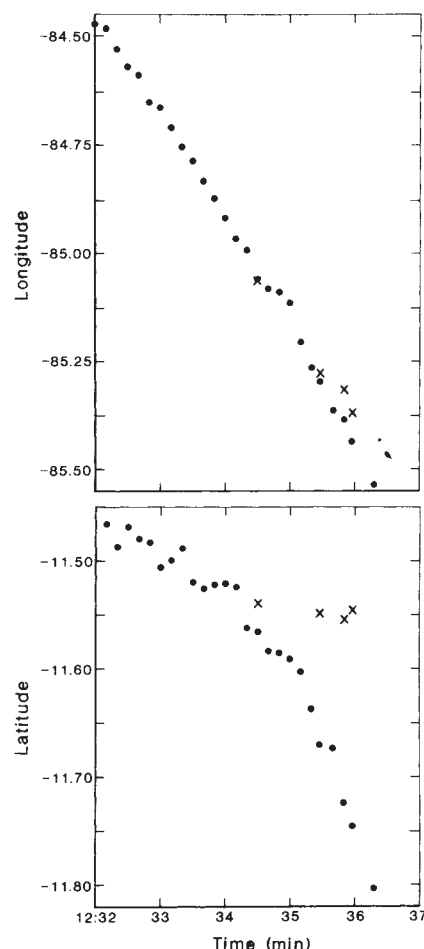
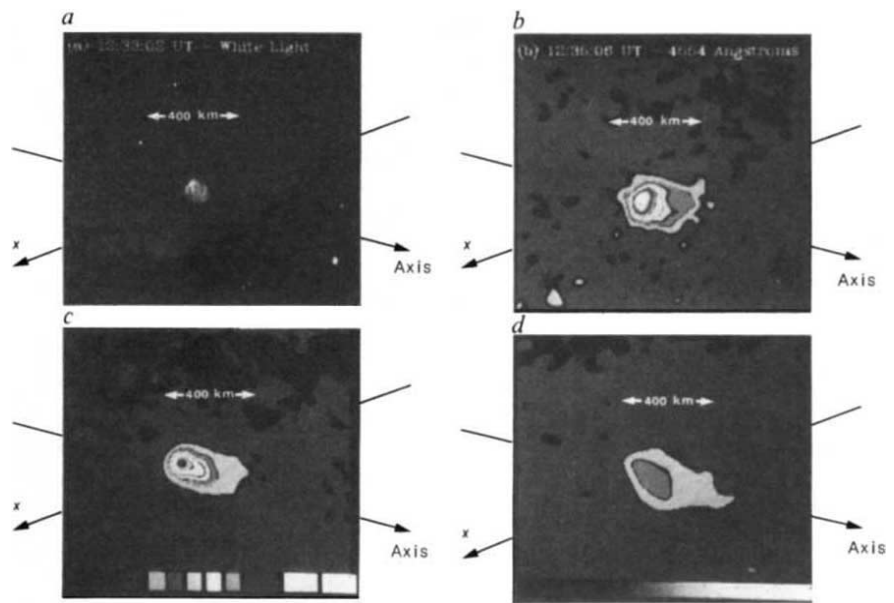


Fig. 1 The apparent motion of the artificial comet in longitude and latitude from the Convair 990 LLLTV observations. The barium gas, was released at 12:32:01 universal time (UT). The motion in longitude is due to the rotation of the earth. The motion in latitude is inertial until 12:34 UT, after which there is a southward acceleration to a maximum final velocity component (perpendicular to the line of sight) of 5.4 km s^{-1} . •, Ions; ×, neutral particles.

Fig. 2 Images of the head of the artificial comet obtained from the CCD (charge-coupled device) imager at Boulder Canyon, Colorado at *a*, 12:33:02 UT (white light) and at *b*, 12:35:08; *c*, 12:35:43; and *d*, 12:36:31 UT (Ba^+ emission line at 4,554 Å). The images show the skewed orientation of the head of the comet with respect to the sun-comet vector (marked *x*). The ellipsoidal core (with long axis of symmetry marked 'axis') and coma extension, ~35–40° southward of the *x* axis, are well shown in *b* and *c*.



During the initial phase (0–2 min), there was no optical evidence for motion of the comet head in the direction of the solar wind. The DIS and LLLTV images showed that a faint, curved ion tail at least 3,000 km long developed within the first 90 s. The DIS data also showed that the intense ion core was surrounded by a much weaker and asymmetric ion halo. About 90 s after the release, the outermost contours detectable by the DIS defined a halo of diameter 500 km, with a sharp southward boundary and a diffuse northward boundary.

Bulk motions of the neutral and ion clouds. The goals of the artificial comet experiment centred on the study of acceleration mechanisms for an ion cloud introduced into the solar wind. It had been anticipated that the dense plasma core would start to move slowly in the direction of the streaming solar wind, eventually reaching a speed comparable to that of the solar wind, in ~30–45 min². Observing plans were formulated on the basis of such expectations; thus, the demise of the comet in slightly more than 5 min has revealed mechanisms not considered in the original formulation of the experiment. The extraction of barium ions, accelerated from the core by the electric field created by the convection of solar-wind plasma and the IMF past the comet-like obstacle, provided an unexpected rapid attrition rate of barium ions into the solar wind. These extracted ions form the diffuse halo around the core observed by the DIS, which were also observed in the *in situ* plasma measurements by the IRM⁶.

The imaging results presented here offer unique spatial coverage of the comet's evolution in time, because of the ability to monitor both neutral and electrodynamical processes. As shown from the LLLTV results in Fig. 1, the ion and neutral clouds drifted together for ~120 s. The initial apparent zonal motion of the neutrals could not be distinguished from that of the IRM, which was westward in the terrestrial reference frame because of the rotation of the Earth. However, the initial southward velocity of the neutral cloud (~0.7 km s⁻¹) exceeded that of the IRM, presumably due to the dynamics of the barium release process.

Between 12:34 and 12:36 UT, there was a surprising motion of the ion cloud core with respect to the neutrals. Rather than seeing a separation in the antisolar direction, as solar wind/plasma cloud interactions accelerated the ions away from the sun, the ion cloud core accelerated southward with respect to the neutrals, reaching a constant velocity of 5–6 km s⁻¹ by 12:35 UT. After about 12:36 UT (+4 min), there was a noticeable acceleration of the depleted core in the direction of the solar wind. This antisunward motion of the plasma after 12:36 con-

formed with the 'late-time' morphology and dynamics patterns predicted before the experiment².

The onset of the antisunward acceleration of the core coincided with a rapid decrease in brightness of the comet as plasma was withdrawn from the head. The disintegration of the head and final demise of the comet occurred during the fifth minute of observations. As discussed in detail in ref. 6, the observed southward acceleration of the core balanced the direct extraction of barium ions from the ion cloud core into the solar wind by the approximately northward-directed electric field. The antisunward acceleration during the first 4 min acted by accelerating individual ions into the tail from the antisunward side of the head, and did not produce a bulk acceleration of the core.

Fine structure and evolution of the artificial comet. Figure 2 shows high-spatial-resolution pictures of the head of the artificial comet obtained from the CCD (charge-coupled-device) imager at Boulder Canyon, Colorado. The head contained an intense ion core that was gibbous-shaped in cross section (Fig. 2*b*), located at the northern edge of the head. The head itself was ellipsoidal with the axis of symmetry ~30–40° southward of the solar-wind direction (–*x* axis; Fig. 2*c*, *d*). The direction of this axis may correspond to the initial path of ejected tail rays described below. The unexpected elongation of the head itself resulted from a combination of solar-wind-aligned erosion and the acceleration of the core transverse to the solar wind, as shown in Fig. 1. The images in Fig. 2 are distinctly non-comet-like in that no true tail features can be seen; the visible elongations are totally within the head of the comet and, moreover, are not in the direction of the solar wind. High thin clouds in the Boulder area prevented the observation of the faint solar-wind-aligned structures from that site.

Figure 3 shows both the head of the artificial comet and a 4,000-km length of the tail, as observed by the LLLTV on the Convair 990. The head contains an intense ellipsoidal core that corresponds to the entire image shown in Fig. 2*b*. The projected Geocentric, Solar Ecliptic (GSE) coordinate axes are shown (*x* towards the Sun, *y* eastward and *z* toward the ecliptic north pole). The mean IMF (12:28–12:32 UT) was directed mainly toward the observer, with smaller components in the southward and antisunward directions ($B_x = +9.5$ nT, $B_y = 5$ nT, $B_z = -1.6$ nT). The mean solar wind velocity vector before the release was close to the –*x* direction shown in these images (5° toward the *z* axis, 3° toward the *y* axis).

The images in Fig. 3, taken 2 s apart, show the apparent rapid evolution of a small feature which seemed to move ~1,200 km

antisunward between the two frames (white arrows). This, and several similar features, accelerated to an apparent antisunward velocity of $\sim 600 \text{ km s}^{-1}$ within 4 s of leaving the boundary of the comet head. These intermittent jets of plasma are directed $\sim 20^\circ$ south of the antisunward direction. Together with the antisunward flow, the ejected plasma gives a curved appearance to the tail in some of the images.

The development of the head and tail of the comet is also shown in the sequence of four images from the Doppler Imaging System (DIS) on the Convair 990 (Figs 4 and 5). These figures show a faint diffuse tail component which curved northward, as well as a 'straight' tail component which probably corresponds to the classical narrow ion tail observed in natural comets (see also Fig. 3). A third, wide-angle, imaging system on the Convair 990 followed the cometary tail to a distance of 15,000 km in the anti-sunward direction, limited only by the field of view of the imager.

Ion velocity distributions in the coma and tail. The DIS results in Fig. 4 were obtained using the 4,554-Å barium ion emission line imaged through a Fabry-Perot interferometer. This technique yields the spatial morphologies observable by any imaging system, and also details of the velocity distribution (or derived temperatures) of the emitting gas.

The results in Fig. 4 do not show a modulation by the Fabry-Perot fringe pattern. This indicates that, within the coma and tail of the comet, the barium ions have a random line-of-sight velocity dispersion of $\geq 34 \text{ km s}^{-1}$. A simple bulk line-of-sight velocity would have caused a mean Doppler shift, not the loss of fringe contrast observed in Fig. 4. To account for this Doppler broadening by a thermal velocity distribution, an ion temperature of $> 10^6 \text{ K}$ (or 200 eV per ion) is required. Part of this velocity dispersion could have been due to initial heating of the ions by plasma wave turbulence, followed by acceleration of the ions by the broadly northward electric field⁶. In the immediate vicinity of the comet-like obstacle the solar wind and electric field will be of reduced magnitude, and the flow will be disordered by mass loading due to the extracted barium ions.

In order to obtain information regarding low-velocity (cold) ion cloud patterns, the DIS images in Fig. 4 were re-processed; the result is shown in Fig. 5. Here one can see an intense ion core within the head which is elliptical in cross-section, and $\sim 160 \times 100 \text{ km}$ in size. This dense core stands out only in the images at 12:34, 12:35 and 12:36 UT (Fig. 5*b-d*), when the core overlies the central 'rest fringe' (innermost dashed line) of the circular Fabry-Perot patterns. The core thus consists predominantly of cold barium ions, and its shape, location and orientation coincide with the intense gibbous and elliptical features observed in the CCD and LLLTV images shown in Figs 2 and 3. As in those figures, the core coincided initially with the location of the diamagnetic cavity, as identified by the IRM. Its long axis was aligned at between 20 and 45° to the GSE z axis, with the southern edge pointed upstream into the solar wind (compare Fig. 2). By 12:36 UT (Fig. 5*d*) the central core had decreased in intensity and size, appearing as a circular feature $\sim 100 \text{ km}$ in diameter.

The DIS fringes allow us to estimate the temperature of the core ions. This temperature was $\leq 5,000 \text{ K}$, in marked contrast to the strongly heated or accelerated plasma in the extended regions of the diffuse coma and tail. The sunward transition between the cold plasma in the core and the hot plasma of the coma was very abrupt (one pixel in the DIS images, or $\sim 30 \text{ km}$). Figure 5 also suggests that within the southern part of the elongated head, there was a region $\sim 200 \times 200 \text{ km}$ in size which may have contained barium plasma with an intermediate temperature or velocity dispersion. This region corresponds to the most antisunward contours visible in Fig. 2, and was the origin of both the straight and curved ion tail components visible in Figs 3 and 4. A core temperature of the order of 5,000 K is consistent with the barium production from the thermite release, the subsequent photoionization, and the ion heating processes

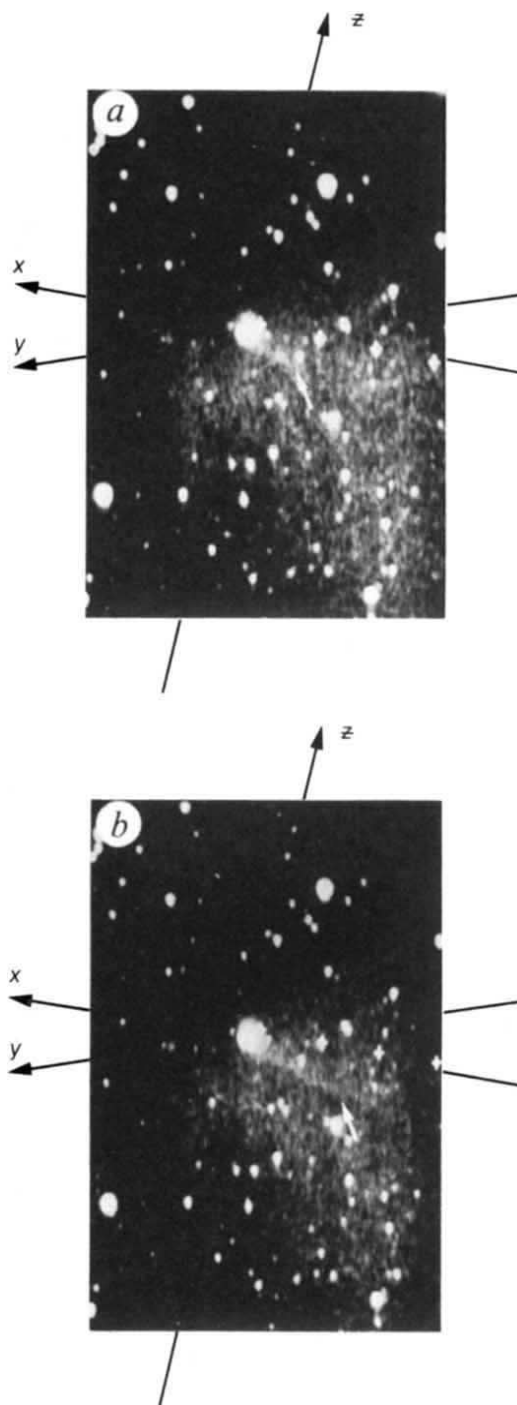


Fig. 3 Images of the artificial comet at *a*, 12:34:53 and *b*, 12:34:55 UT, obtained from the LLLTV on the Convair 990. The form of the extended tail can be seen, and a weak feature (marked by white arrow) is seen to propagate over $\sim 500 \text{ km}$ in the 2-s interval between the two images. Several similar weak and rapidly moving tail features were observed between 12:33 and 12:35 UT. The projections onto the image plane of the GSE coordinate system (x toward the Sun, y eastward, z toward the ecliptic north pole) are indicated.

which may have occurred when the initial diamagnetic collapsed, and the compressed IMF penetrated the core ion cloud.

Conclusions

Combined airborne and ground-based optical diagnostic systems have been successful in observing the AMPTE artificial comet experiment of December 1984. The plasma cloud intro-

Fig. 4 Sequence of four images of the comet obtained by the DIS on the Convair 990, at intervals of ~ 50 s. There is a diffuse tail, originating $\sim 15^\circ$ south of the antisunward ($-x$) direction, and curving northward toward the x axis. A straight tail, $\sim 20^\circ$ south of the ecliptic (x - y) plane, is also visible. This straight tail matches the direction of propagation of the faint feature depicted in Fig. 3. The diffuse coma and tail regions do not display Fabry-Perot fringe structures (dashed lines) which would correspond to the fringes generated by cold barium plasma, indicating a line-of-sight velocity dispersion of at least 34 km s^{-1} within these regions.

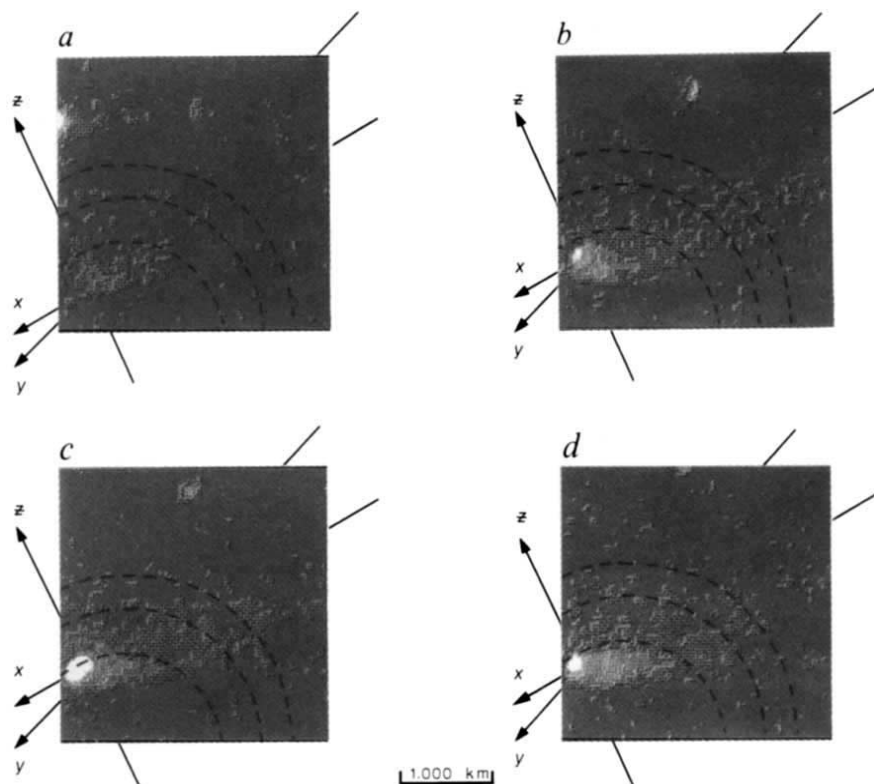
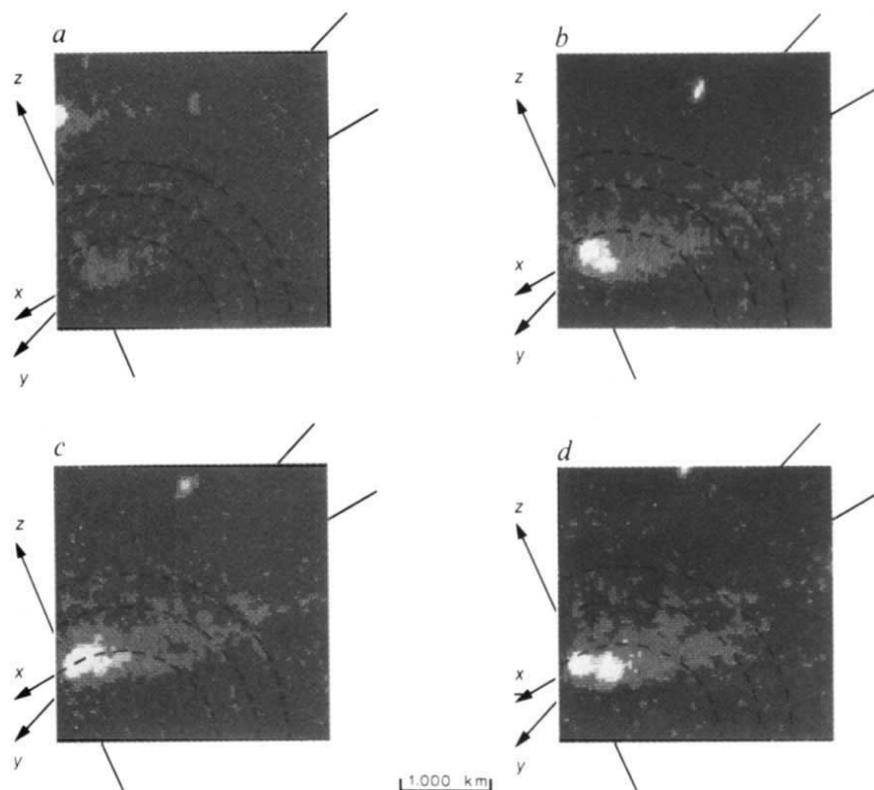


Fig. 5 The image data of Fig. 4 re-displayed with a different intensity scaling. This highlights a small, intense region of cold barium lying within the extended diffuse coma. This region is highlighted in the DIS images only when it coincides with a 'rest' fringe of the Fabry-Perot interferometer (first fringe). The Doppler-shift of the core of cold barium is small, indicating a line-of-sight velocity component of $< 5 \text{ km s}^{-1}$.

duced into the solar wind was surprisingly short-lived (~ 5 min), as a result of rapid extraction of barium ions by the electric field induced by the solar wind and IMF. The imaging results presented here revealed that the head became extended along an axis 35 – 40° from the direction of solar wind flow. Tail-like features appeared occasionally in impulsive bursts, extending in the solar wind direction. Doppler measurements showed that the head was composed of a central core of cold ions, whereas

the extended coma and tail were formed by more energetic ions. The absence of an initial acceleration of the core in the antisunward direction, coupled with an observed drift transverse to the solar wind, point to electrodynamic processes not considered previously⁶. Momentum balance with the extracted ions caused the southward acceleration, while the tailward acceleration appeared to be concentrated in a small part of the ion cloud on the antisunward side of the head, rather than causing a bulk

antisunward acceleration of the entire head and core. Confirmation of the transverse motion of the head with respect to the solar wind direction was obtained by the ground-based CCD imaging system during the second artificial comet experiment of July 1985⁷.

The airborne observations were made possible by the technical

and engineering assistance of the Medium Altitude Missions Branch of NASA Ames Research Center, under the direction of Bob Cameron. The Boston University observations were funded by the Max Planck Institute and NASA. The University of Alaska observations were supported by NASA. The UCL involvement was sponsored and partly funded by SERC.

Received 2 December 1985; accepted 2 March 1986.

1. Krimigis, S. M., Haerendel, G., McEntire, R. W., Paschmann, G. & Bryant, D. A. *Eos* **63**, 843–850 (1982).
2. Haerendel, G. *Eur. Space Ag. J.* **7**, 135–144 (1983).

3. Rees, D. *et al. Planet. Space Sci.* **32**, 273–285 (1984).
4. Valenzuela, A. *et al. Nature* **320**, 700–703 (1986).
5. Lühr, H. *et al. Nature* **320**, 708–711 (1986).
6. Haerendel, G., Paschmann, G., Baumjohann, W. & Carlson, C. W. *Nature* **320**, 720–723 (1986).
7. Enriquez, J., Mendillo, M., Baumgardner, J. & Haerendel, G. *Eos* **66**, 1032 (1985).

In situ magnetic field observations of the AMPTE artificial comet

H. Lühr*, D. J. Southwood†, N. Klöcker*, M. W. Dunlop†, W. A. C. Mier-Jedrzejowicz†, R. P. Rijnbeek†, M. Six†‡, B. Häusler‡ & M. Acuña§

* Institut für Geophysik und Meteorologie der Technischen Universität Carolo-Wilhelmina, 3300 Braunschweig, FRG

† Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

‡ Max-Planck-Institut für Physik und Astrophysik, Institut für Extraterrestrische Physik, 8046 Garching, FRG

§ NASA, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Magnetometers aboard the IRM and UKS spacecraft monitored the magnetic field during the AMPTE artificial comet experiment of 27 December 1984. Rapid photoionization of the released barium vapour resulted in the formation of a magnetic cavity, shielded from the ambient magnetic field. The presence of this highly conductive obstacle caused draping and compression of the solar wind magnetic field.

ON 27 December 1984, the release of barium from two canisters carried by the West German IRM (Ion Release Module) spacecraft marked the creation of the first artificial comet^{1–3}. As part of the active phase of the AMPTE (Active Magnetospheric Particle Tracer Explorers) mission, ~1 kg of barium ions was released in the solar wind on the dawn flank of the Earth's magnetosphere. The high ionization rate (2.4×10^{23} ions s⁻¹) in the initial phase and the heavy mass of the barium (137.4 AMU) resulted in the formation of an ion cloud dense enough to exclude the interplanetary magnetic field and disrupt the solar wind flow^{2,3}.

The magnetic field instruments on both spacecraft are fluxgate magnetometers, and are described in detail elsewhere^{4,5}. Both instruments are capable of determining the ambient field to a relative accuracy of <0.1 nT. An intercalibration of both magnetometers performed in the solar wind before the release ensured correct offset and sensitivity determination.

At the time of the release both the IRM and the UKS (United Kingdom Satellite) were located in the solar wind close to the ecliptic and on the dawn flank of the magnetosphere. The spacecraft were separated by 170 km; in terms of the solar wind flow direction, the UKS was located downstream of the IRM. The angle between the spacecraft separation vector and Sun-Earth line in the ecliptic plane was ~60°, as sketched in Fig. 1; however, uncertainties in the orbit calculations result in an uncertainty of several tens of degrees in this angle. The spacecraft locations were chosen so that the two magnetometers might record very different perturbations. In what follows we describe the ambient field conditions before and after the release as monitored by the IRM, and then give a description of the interior

of the magnetic cavity and the field distribution around the artificial comet.

Ambient magnetic field conditions

The barium release was intended to occur in the magnetosheath region on the dawn flank of the magnetosphere¹. At the time of the release the UKS and the IRM were at $x = 4.9 R_E$, $y = -16.5 R_E$, $z = -0.9 R_E$ (where R_E is the radius of the Earth) in geocentric solar ecliptic (GSE) coordinates (x towards the Sun, y eastward and z towards the ecliptic north pole). Both spacecraft were located inside the average location of the terrestrial bow shock as determined by Fairfield⁶. However, the shock

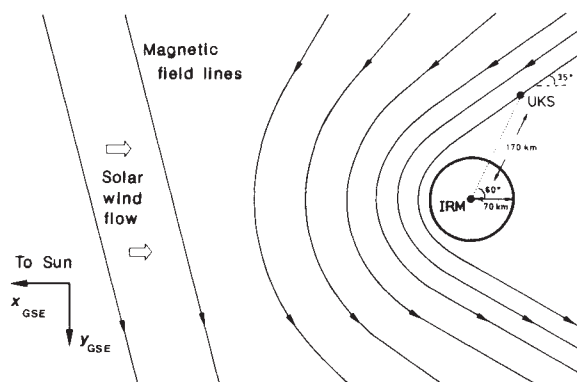


Fig. 1 Simplified sketch of the artificial comet showing the interplanetary magnetic field lines draped over the released Ba⁺ ion cloud (shaded). The initial positions of the IRM and UKS relative to the comet are also indicated. The Ba⁺ cloud moves in response to the force exerted by the solar wind, leaving both spacecraft in the region outside the comet.

‡ Permanent address: Institut de Physique du Globe, 4 Place Jussieu, 75230 Paris Cedex 05, France.

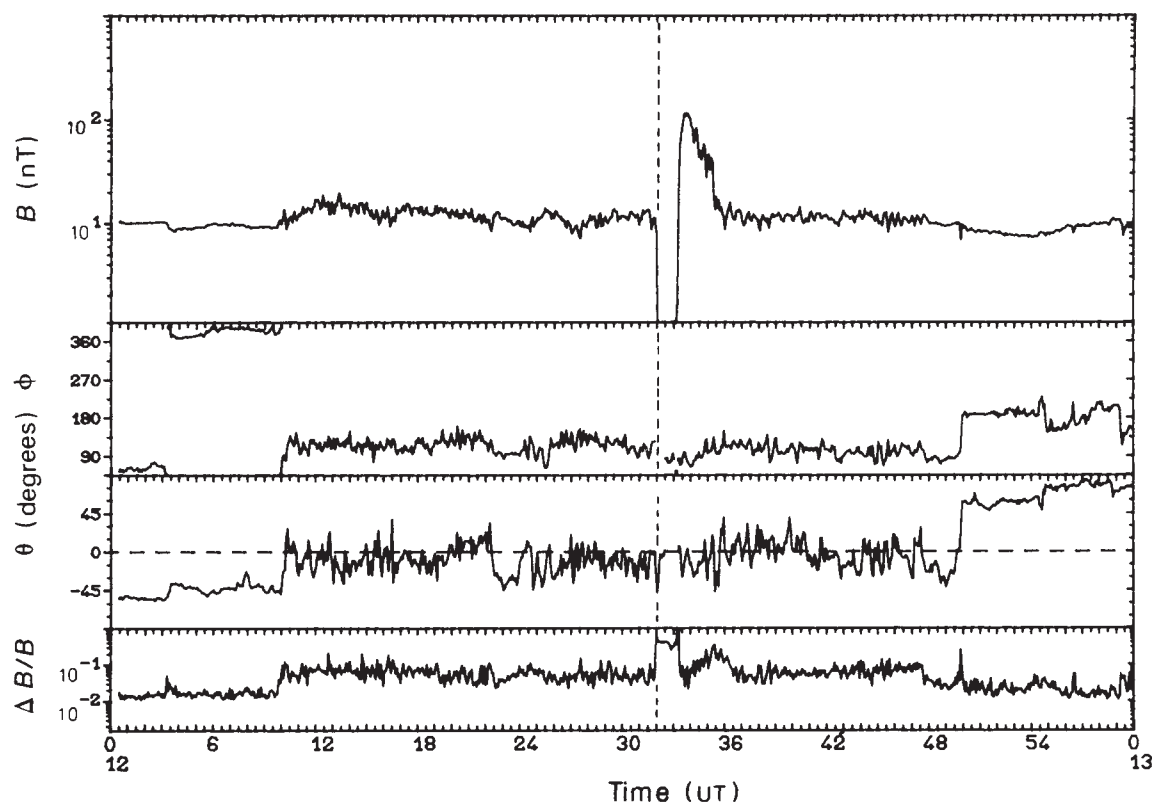


Fig. 2 One hour of IRM magnetic field data (4-s averages), showing the ambient field conditions during the artificial comet release on 27 December 1984. The time of release (12:32 UT) is indicated by the dashed vertical line. The panels show (from top to bottom): field strength, B ; azimuth, ϕ , and elevation, θ , in geocentric solar ecliptic (GSE) coordinates; and the variability of the field, $\Delta B/B$, over one spin period. The release event is embedded in an area of foreshock wave activity, which is a common feature for magnetic field lines connected to the Earth's bow shock.

position is highly variable, and the high solar wind speed on 27 December led to the shock being swept anti-sunward of the IRM and UKS 40 min before the release. The release thus took place in the solar wind.

Because the separation of the UKS and the IRM was only 170 km, we may discuss the ambient field conditions from the records of a single spacecraft. Figure 2 shows the background magnetic field as recorded by the IRM magnetometer during the hour centred on the release event. The data are presented in GSE spherical coordinates, showing from top to bottom the magnitude, B , field-line direction azimuth, ϕ , relative to the Earth-Sun line ($\phi = 0$ is sunward) and elevation, θ , above the ecliptic, and the field variability, $\Delta B/B$. Perturbations attributed to the release start at 12:32 UT (vertical dashed line) and continue for 4 min.

The solar-wind magnetic field is quiet until 12:10 UT. After this time, the measured direction of the field, $\phi \approx 120^\circ$, $\theta \approx 15^\circ$, implies that the field is connected to the Earth's bow shock earthward of the spacecraft. This inference is supported by the presence of large fluctuations in field amplitude and direction due to foreshock waves generated by ions reflected from the shock. Plasma instruments on both spacecraft confirmed the presence of such ions (G. Paschmann and A. D. Johnstone, personal communication).

At $\sim 12:50$ UT, 18 min after the release, the field lines swung away from the spiral configuration to an anti-sunward direction with a large positive elevation (Fig. 2). With that the spacecraft were back in a magnetically quiet environment.

The release of barium ions thus took place just sunward of the bow shock, but on field lines connected to it. For this reason,

the field imposed by the solar wind during the release was highly variable. The strength was ~ 10 nT and the mean direction was $\sim 120^\circ$ in azimuth, with only slight elevation out of the ecliptic ($\sim -10^\circ$). The major magnetic field component was therefore at right angles to the ambient solar wind flow.

Characteristics of the magnetic cavity

The two barium-filled canisters were deployed from the IRM at 12:21:59.86 UT. At the time of ignition, 600 s later, the two release canisters were 1.75 km apart, with the IRM halfway between them. The two expanding ion clouds merged as they encountered the IRM at 12:32:00.72 UT. Figure 3 shows the field perturbations measured by the UKS (thin line) and the IRM (bold line) during the release. The most striking feature is the complete absence of magnetic field at the IRM in the first stages. Before entering the cavity, the IRM measured a transient field enhancement of a factor of two. Thereafter the field strength decreased exponentially, with a time constant of 0.3 s.

Figure 4 shows in more detail the field distribution inside the magnetic cavity as recorded by the IRM. The measurements of the three components have been averaged over 4.4 s (1 spin period) to eliminate residual spin signal in the data and to achieve a relative resolution of 0.02 nT. We conclude that between 12:32:12 and 12:32:30 UT the field strength remained below 0.02 nT; the external field was thus shielded by a factor of > 500 . As time progressed the shielding became less effective. The components transverse to the solar wind, B_y and B_z , increased slowly but significantly while B_x remained close to zero for ~ 1 min. About 75 s after the Ba^+ release, the IRM left the cavity. The field strength increased exponentially, with a

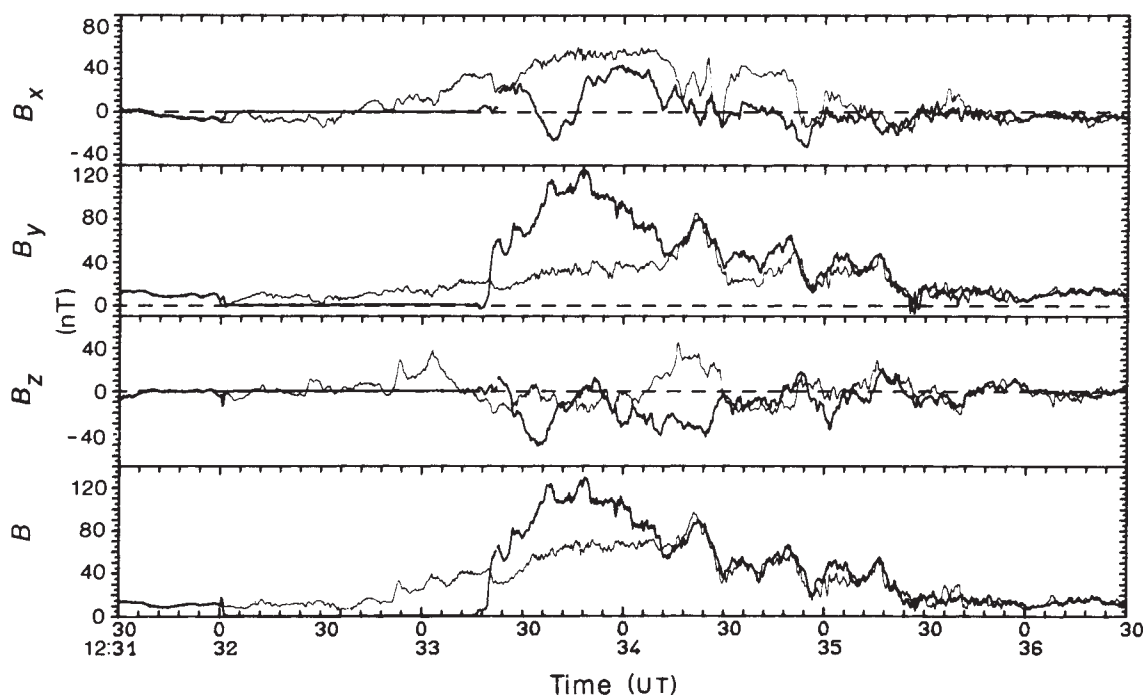


Fig. 3 Five minutes of combined UKS (—) and IRM (---) magnetic field data (0.25-s averages), near the time of barium release (12:32 UT). The data are plotted in GSE cartesian coordinates (x sunward, y eastward, z toward the ecliptic north pole).

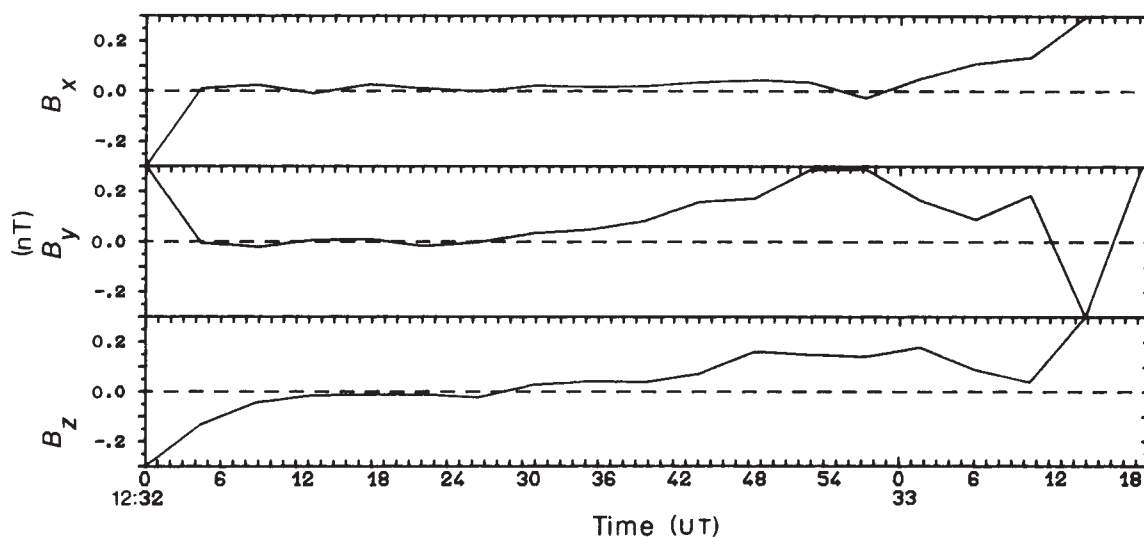


Fig. 4 Residual magnetic field inside the magnetic cavity. The IRM data have been averaged over one spin period (4.4 s) to achieve a resolution of 0.02 nT. Until 12:32:30 UT the field is below the sensitivity threshold of the instrument, thereafter B_y and B_z , the components perpendicular to the solar wind flow, rise slowly whilst B_x remains close to zero. At 12:33:15 UT IRM exits the cavity.

time constant similar to that which characterized the decrease on entering the cavity.

Field distribution outside the cavity

Figure 3 shows that while the IRM was inside the cavity, the UKS monitored a gradual increase in the exterior field which started ~ 35 s after the release, when the cavity radius was ~ 30 km. Figure 5 shows the fields at the IRM and UKS projected in the plane of the ecliptic. The increase in field strength at the UKS is marked also by a rotation of the field towards the solar

(x) direction. The direction change is clearly associated with the presence of the cavity, for when the IRM exits it encounters a field with a very different direction (mainly transverse to the solar wind). The radius of the cavity is likely to be ~ 70 km at the time the IRM exits, and the obvious explanation of the radical directional difference in the measured fields is that the field is draped over the obstacle.

The field strength encountered by the IRM immediately adjacent to the cavity is greater than that measured at the UKS, and it continues to increase at a greater rate than at the UKS until

it reaches an approximate plateau near 12:33:40 UT (Fig. 3). The peak strength measured is 130 nT, 12 times the field that is believed to be present upstream in the solar wind. At the same time, the field strength at the UKS stabilizes at a value of ~ 70 nT.

At $\sim 12:34$ UT the field starts to decrease at the IRM, but not at the UKS. After 12:34:10 UT the field strengths measured by both spacecraft are identical. Both spacecraft then experience a series of pulses in field strength superposed on a decreasing background. The field strengths have returned to solar-wind values by 12:35:30 UT; however, until 12:35 UT the directions are substantially different (Fig. 5). The UKS systematically records a larger B_x (sunward) component, as shown in Figs 3 and 5.

Discussion

In highly conducting media such as both the barium ion cloud and the ambient solar wind, the magnetic field plays a fundamental role in the distribution of stress. The body force $\mathbf{J} \times \mathbf{B}$ (where $\mathbf{J}$ is the current density vector) can be described by a magnetic stress tensor. The field exerts a pressure $B^2/2\mu_0$ (where μ_0 is the magnetic permeability of free space) and there is a tension B^2/μ_0 along the field. In the present experiment, the compression of the magnetic field exerts a pressure on the cavity, but the skewing of the field in the solar direction recorded at the UKS also indicates a tensile force parallel to the solar wind flow. The 35° tilt in field in the ecliptic plane recorded by the UKS in the first half of the event (Fig. 5) is consistent with the notion of the field at the UKS draping about an obstacle of ~ 70 km radius, given the nominal relative positions of the UKS and IRM. The UKS is clearly on the flank of the cavity throughout the interaction.

Until 12:34 UT the UKS records a fairly steady field-line draping. In the latter phases it appears that at both spacecraft the degree of draping increases and decreases, indicating that the net tension exerted on the obstacle accumulates and is released in the three major pulses shown in Fig. 3. However, as the lowest panel of Fig. 3 shows, the field pressure is also pulsing. This pressure force is roughly in anti-phase with the skewing of the field.

It is important to note that the magnetic tensile force on the obstacle deduced from our data is in the solar wind direction. The field acts as the medium for momentum transfer from the streaming solar wind to the obstacle formed by the cavity. However, one must not adopt too simple a view. The obstacle is unlike a planetary obstacle (such as that formed by the Earth's magnetosphere) because it is very small compared with either the solar-wind or barium-ion gyroradius (>500 km). The ions cannot therefore be considered to be frozen to the field lines, as in a simple magnetohydrodynamic model. The electron gyroradius is small (~ 300 m for solar wind electron), so they can be frozen in. Thus the electron fluid acts as an intermediary in the momentum transfer process.

The role played by the electrons is evident in the thin current sheets encountered by the IRM at the edge of the cavity. If we assume that on entry the boundary passes the IRM at the expected Ba^+ cloud expansion speed of $\sim 1 \text{ km s}^{-1}$, the current-sheet scale is 300 m. This is so small that the current must be carried by cold photoelectrons from within the cloud. It is interesting to note that on exit the observed timescale for field strength increase was the same. If the same particles are responsible for the current structure, then the exit velocity was probably

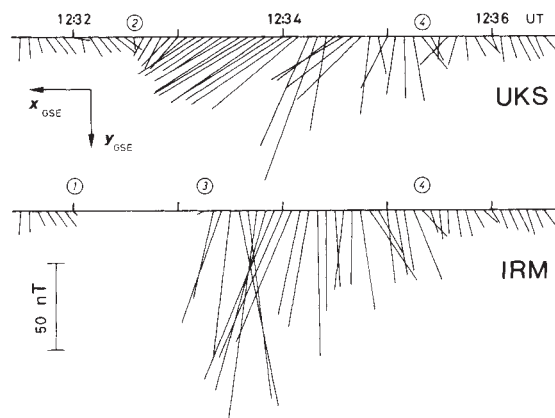


Fig. 5 Projection on to the GSE x - y (ecliptic) plane of magnetic field vectors (5-s averages) measured at the UKS and IRM during the barium release. Major events are labelled as follows: ①, release of barium field ions; ②, encounter of UKS with draped field lines; ③, IRM exits the magnetic cavity; ④, simultaneous return to solar wind field conditions at both UKS and IRM.

also on the order of a few km s^{-1} . The resulting current density on entry is $20 \mu\text{A m}^{-2}$.

The trajectory of the IRM's exit from the cavity is as yet uncertain. The present interpretation of the images suggests that there was a substantial southward component in the cloud motion at the time. The absence of draping at the IRM, in the field immediately outside the cavity, suggests that the exit was made in the X - Z plane, indicating that there was little east-west cloud motion^{2,3}.

The magnetic perturbations recorded during the AMPTE lithium ion releases⁷ bear a resemblance to those reported here, and particularly to those seen during the release made in very similar upstream foreshock conditions on 11 September 1984. The amplitude and timescale are of course radically different. The AMPTE Li^+ releases have been described as 'cometary'⁸. It is important to ask what relevance the data presented here have to cometary behaviour. The ion tails of comets are believed to result from the interaction of the solar wind magnetic field with the cometary ionized environment, causing draping of the field lines around the obstacle as the solar wind flows by. Undoubtedly, we have detected similar trapping of the field by the artificial comet, as illustrated in the configuration of the field lines around the cavity shown in Fig. 1. However, the cometary ion tail is believed to result from ion motion along the field away from the comet head; and there is no evidence for such behaviour in this experiment. In this respect the magnetic interaction we have seen may not be cometary. What renders this experiment unique in the study of solar wind-obstacle interactions is the small size of the obstacle and thus the unmagnetized behaviour of ions. What is most important is that the field stresses are still seen to act on a gross scale in the manner expected for larger obstacles. This remains the most important conclusion from this first review of the magnetic data.

We thank H. Ruprecht and M. Lampe for their support in evaluating the IRM magnetometer data. The research at the Technical University of Braunschweig was supported by grants of the Bundesministerium für Forschung und Technologie. The UKS programme is supported by SERC.

Received 2 December 1985; accepted 18 March 1986.

1. Krimigis, S. M., Haerendel, G., McEntire, R. W., Paschmann, G. & Bryant, D. A. *Eos* **63**, 843 (1982).
2. Valenzuela, A. *et al.* *Nature* **320**, 700-703 (1986).
3. Haerendel, G., Paschmann, G., Baumjohann, W. & Carlson, C. W. *Nature* **320**, 720-723 (1986).

4. Lühr, H., Klöcker, N., Oelschlägel, W., Acuña, M. & Häusler, B. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 259 (1985).
5. Southwood, D. J., Mier-Jedrzejowicz, W. A. C. & Russell, C. T. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 301 (1985).
6. Fairfield, D. H. *J. geophys. Res.* **76**, 6700 (1971).
7. Lühr, H. *et al.* *J. geophys. Res.* (submitted).
8. Southwood, D. J. & Bryant, D. A. *Nature* **312**, 594 (1984).

UKS plasma measurements near the AMPTE artificial comet

D. J. Rodgers*, A. J. Coates*, A. D. Johnstone*, M. F. Smith*, D. A. Bryant,[†]
D. S. Hall[†] & C. P. Chaloner[†]

* Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking RH5 6NT, UK

[†] Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

UKS observations of the ionized barium and disturbed natural plasmas 170 km from the centre of the artificial comet release of 27 December 1984 lasted over 4 minutes. During this disturbance, solar-wind ions were retarded and deflected southwards and dawnwards, while barium was accelerated northwards. Substantial electron energization was seen throughout, varying with time and with the energy of the source electrons. Rapid electron and ion density changes were observed in later stages.

DURING the AMPTE (Active Magnetospheric Particle Tracer Explorers) mission barium release of 27 December 1984, the ion and electron detectors aboard the United Kingdom Satellite (UKS) measured positively charged particles in the energy range 10 eV–20 keV and negatively charged particles between 12 eV and 18 keV. They were able to observe the effect of the release on the ambient solar wind and also, in the latter stages of the release, to monitor the release ions as they were accelerated past the spacecraft.

Unlike the Ion Release Module (IRM), which was in the heart of the diamagnetic cavity created by the release of barium ions, the magnetic measurements made by the UKS show that it never entered the diamagnetic region¹. The magnetic field at the UKS, 170 km away from the IRM, increased steadily until after the IRM had emerged from the cavity. This, combined with the fact that the UKS was downstream of the IRM and displaced from the IRM–Sun line by 60° (ref. 1), suggests that the UKS was on the dawn flank of the obstacle formed in the solar wind by the expanding release cloud.

The characteristic photoionization time of the barium was 28 s (ref. 2); therefore, most of the barium was ionized before the expanding cloud reached the UKS. An isolated ion, created in the solar wind, is accelerated by the interplanetary electric and magnetic fields and develops a cycloidal trajectory. Even in the strongest magnetic fields occurring during the event (~100 nT), the gyroperiod of a barium ion is as long as 2 min and its gyroradius is ~6,000 km. This is much larger than the cavity radius (<170 km) and the distance to which the neutral barium could have expanded in the few minutes that the interaction was observable. We conclude that gyration is unimportant in any of the detected barium ions, and that they have followed essentially straight-line orbits along whatever electric field exists near the spacecraft. In a known electric field, the energy of the detected particle is a direct measure of the distance of the spacecraft from the point of creation. In contrast, the gyroradius of the solar-wind ions ranges from 40 to 400 km, much smaller than that of the barium ions and comparable with the dimensions of the barium cloud.

Ions carry the bulk of the solar-wind momentum. The electrons carry little momentum but are strongly magnetized throughout the event, and their low inertia renders them more readily able to respond to electric fields and to wave-heating processes in particular. Our analysis suggests that electron heating does indeed take place in the vicinity of the release.

Instrumentation

The UKS included in its suite of instruments an ion and an electron energy spectrometer to measure charged particles over all directions and a wide energy range.

The ion instrument³ was capable of operating in several modes which varied its angular, energy and time resolution. During the barium release, a mode was selected to give high time resolution: a full 3-dimensional distribution each spin (5 s). This allowed 45° resolution in both polar and azimuthal angle and 14% energy resolution ($\Delta E/E$) for each of 60 energy levels between 10 eV and 20 keV.

The electron instrument⁴ could also operate in several modes. For this release study it scanned through 24 energy levels in the range 12 eV–18 keV, with 15% energy resolution, once per second. Thus consecutive measurements at the same energy were separated by 70° in azimuth. The polar angular resolution was 28°.

Observations

In Fig. 3 (p. 715), the data from the ion instrument are displayed as three spectrograms, with time as the abscissa and count rate represented by a colour scale. For each of the panels, the ordinate corresponds to one of the three 'dimensions' of the instrument: energy (top), polar angle θ (middle) and azimuthal angle ϕ (bottom). In each case the counts are summed over the other two dimensions. All angles are in satellite coordinates (less than 1.5° from GSE (geocentric, solar, ecliptic) coordinates). Hence $\theta = 0^\circ$ points roughly along z_{GSE} (toward the ecliptic north pole), $\phi = 0^\circ$ along x_{GSE} (sunward) and $\phi = 90^\circ$ along y_{GSE} (eastward). In the top panel, the solar wind can be seen at the start of the plot. It is an intense band with narrow energy range centred around 1.5 keV. Its direction, as seen in the angular panels, is away from the Sun ($\phi \approx 180^\circ$) and in the satellite spin plane ($\theta = 90^\circ$). About one minute after the release, at 12:32:55 UT, the solar wind starts to become disrupted. It is decelerated (its average energy decreases), becomes more irregular, and is deflected southwards and dawnwards so that it is no longer restricted to the two polar angle bins around the ecliptic and one azimuthal bin in which the undisturbed solar wind lies. After maximum deflection, the solar wind gradually returns to the original direction. Following about four minutes of disturbed flow, the satellite leaves the region of the comet and re-enters the undisturbed solar wind.

The electron data are shown in Fig. 4. The large panel is a spectrogram with electron intensity indicated by a colour scale, time as the abscissa and energy as the ordinate. The lower panel is another spectrogram showing 'variability', which is a measure of how much the intensity at each energy varied with time and angle. Intensity increases were observed over the energy range up to 1 keV throughout the period 12:32:30–12:36:00 UT, the nature of the changes in energy spectrum suggesting energy-dependent electron energization rather than simply density increases. Apart from the period 12:33:00–12:33:15, the

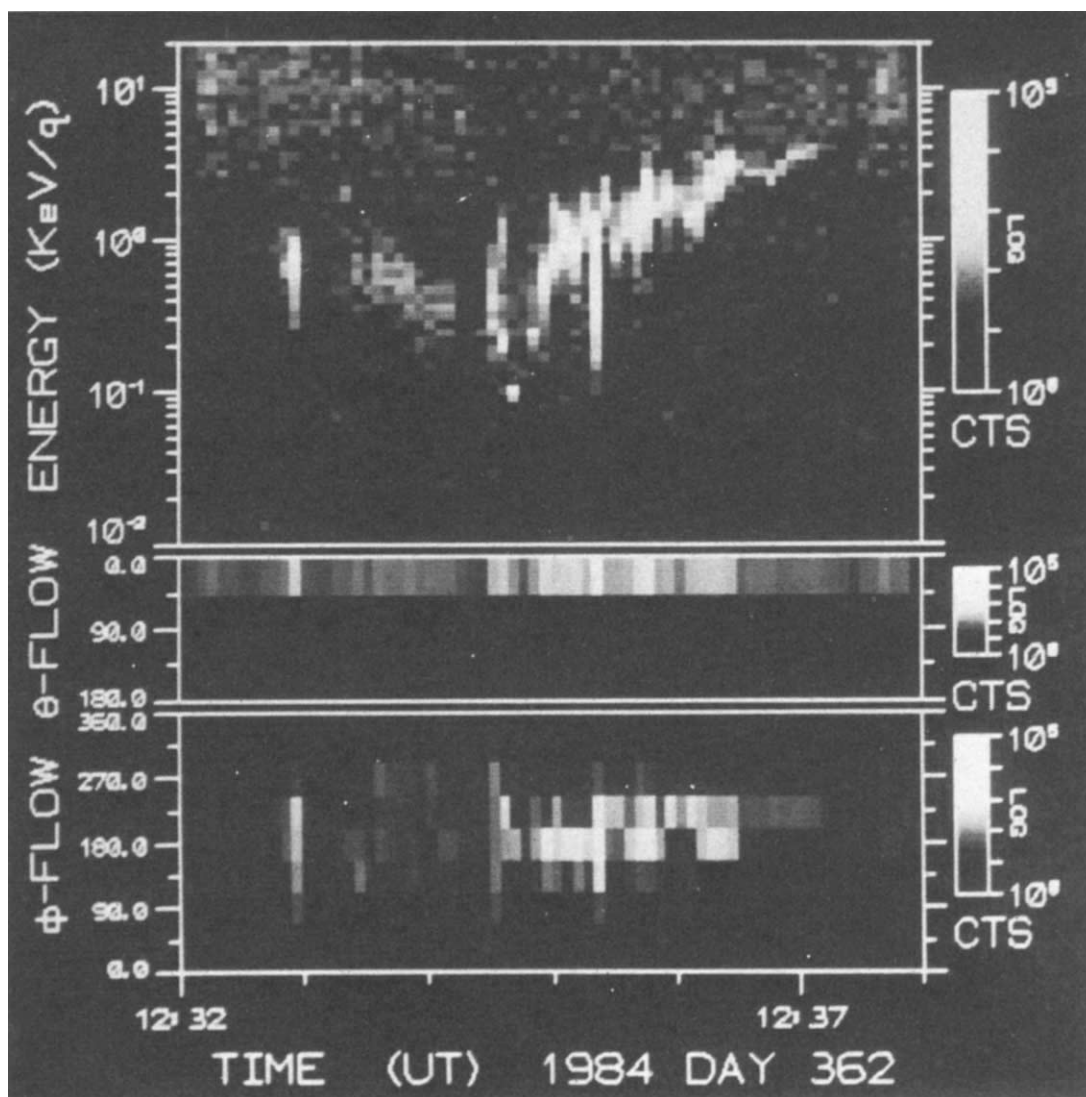


Fig. 1 Ion spectrograms, as shown in Fig. 3, but showing only ions flowing within 45° of north.

intensity is at a relatively low level until there is a sudden increase at 12:34:00, with the highest intensities occurring at 12:34:27. Thereafter, the average energy of the electrons decreases, although not smoothly, until the satellite leaves the disturbed region. Clearly, there are strong variations in the amount of heating throughout. In particular, the heating is most intense when the solar wind is observed to be most retarded.

In the later stages of the encounter (12:34:45–12:36:00), strong, regular fluctuations are observed in the solar-wind ion and electron energy distributions. A gradual increase in the width of the ion distribution is followed by a sudden decrease in distribution width and an increase in the average particle energy. At the same time, the electron intensity shows a brief decrease. This may be evidence of spatial structure, for example, flux tubes caught up by the comet, or of temporal variations, such as a pulsing of the whole comet. The scale size over which the variation takes place must be larger than the separation between the UKS and the IRM, because it is also seen by the IRM⁵.

Analysis

Figure 2 extracts from the data sets of Figs 3 and 4 a number of quantitative parameters. These show the broad behaviour of the plasma, rather than detailed information on the plasma distribution. The top panel (a) shows the energy gained by

electrons that had initial energies of 12 eV (solid line) and 106 eV (dotted line) in the solar wind. This is estimated using the assumption that phase-space density is conserved along their trajectories (Liouville's theorem). The next panel (b) shows electron intensity at 12 and 106 eV; the tick marks labelled 'JSW(12)' and 'JSW(106)' are the undisturbed solar wind intensities.

Figure 2c–e shows the ion data, displayed as bulk velocity, temperature and density obtained by integration over the complete distribution. Because the lower limit of the detector energy range is well above the energy of newly created barium ions, the density shown may be an underestimate. Note also that the experiment mode used at this time has gaps in angular coverage, and so does not give an accurate value of the density in the undisturbed solar wind at the start and end of this period.

Although the release took place at 12:32:01 UT, no change in the solar wind was observed at the UKS until 12:33:10 when there was sudden decrease in bulk velocity from 400 to 300 km s⁻¹ (Fig. 2c). This change in velocity was accompanied by a southward and downward deflection, visible in Fig. 4. This velocity change probably occurred when magnetic field lines that were piled up on the cavity expanded past the satellite. Both ion and electron density increased after this, perhaps as more magnetic flux and its associated plasma accumulated in front of the comet.

At 12:33:52 there was a sudden decrease in ion density (Fig. 2e), while the electron intensities increased even more (Fig. 2b). Because the plasma had to maintain overall neutrality, this imbalance could not have been real; rather, many of the ions were probably unobserved because they lay outside the energy range of the ion instrument. A likely explanation is that the barium cloud had reached the satellite and the undetected ions were barium ions below the threshold energy of the instrument. This is supported by the fact that soon afterwards, at 12:34:25, a population of ions, distinct from the solar wind and almost certainly barium, was detected at the lowest energies of the ion instrument and moving in a southward direction (see Fig. 3). At this time the ion bulk velocity and temperature decreased (Fig. 2c, d), as the cold population became the dominant plasma component.

Simultaneous with the detection of cold barium ions, there was a sharp increase in the electron intensity at energies >100 eV (Fig. 4). The electron energization throughout the disturbed region was energy-dependent, the energy of electrons within the range 37–106 eV being increased by factors of 3–5 while that of electrons of energy below 20 and above 400 eV was increased by factors of less than 2. This energy dependence requires an energy-selective acceleration process such as a resonant interaction with plasma waves⁵. The magnetic field was then at its most intense and predominantly duskward¹, indicating maximum compression of the magnetic field.

After this, the trend in the electron intensity and energy gain and in the ion velocity was back towards the undisturbed levels but all went through several large fluctuations before reaching them (Fig. 2a–c). These fluctuations were closely mirrored by the magnetic field strength¹. The measured ion density was much lower than the electron density throughout this period until it recovered at 12:35:20 (Fig. 2e).

Whereas the magnetic encounter ended at 12:35:30 when the magnetic field reverted to its initial strength and direction¹, the final discernible plasma interaction was at 12:36:20 when the ion temperature peaked (Fig. 2d). This can also be seen in Fig. 3, which shows that there were many ions above the solar-wind energy at this time.

Discussion

When the UKS first entered the dawn flank of the comet at 12:33:10, there was a sharp increase in the electron energy gain. This can also be seen as an increase in intensity in the electron spectrogram (Fig. 4). At the same time there was a peak in ion density (Fig. 2e) and the first major change in the magnetic field¹. It is most likely that this plasma change is associated with an Alfvén wave, which communicates the disturbance in the magnetic field along the field direction, causing the field lines to bend around the comet. Immediately after this first disturbance, the solar wind slowed down and was deflected southwards and dawnwards.

This deflection brought the solar wind antiparallel to the interplanetary electric field as determined from $-\mathbf{v}_s \times \mathbf{B}$, where $\mathbf{v}_s$ is the velocity of the solar wind. This is the direction required to extract both momentum and energy from the flow of solar-wind ions. Once the disturbance starts to build up, the electric field is no longer equal to $-\mathbf{v}_s \times \mathbf{B}$. In fact, $|\mathbf{E}|$ must be $<|\mathbf{v}_s \times \mathbf{B}|$ so that the solar wind is deflected against the electric field by the Lorentz force and thus loses energy. Thereafter, the electric field will also be modified by the charge separation caused by the different behaviour of ions and electrons. Electrons are strongly magnetized and drift perpendicular to the electric field. The ion gyroradius is comparable with or greater than the size of the cloud and the ions move predominantly in the direction of the electric field. The electric field was not measured directly and cannot be deduced from the $\mathbf{v}$ and $\mathbf{B}$ measurements but we assume that, at least near the UKS, it is in the opposite direction to the Lorentz force, if not of the same magnitude.

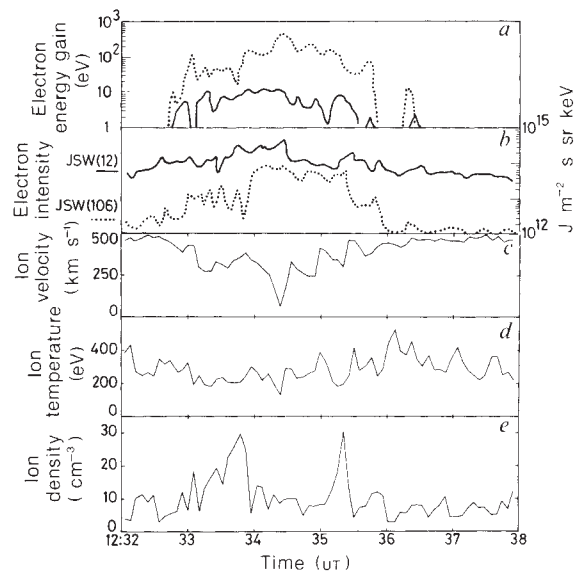


Fig. 2 Plasma bulk parameters plotted against time. *a*, Energy gained by electrons that had initial energies of 12 eV (solid line) and 106 eV (dotted line); *b*, electron intensity at these energies. The tick marks labelled 'JSW(12)' and 'JSW(106)' are the intensities at 12 and 106 eV averaged over the 30 s before 12:32:01 UT (that is, the undisturbed solar wind intensities). *c*–*e*, Ion data.

When the low-energy barium ions are observed, at 12:34:20, they are moving southwards. At this time $\mathbf{v}_s \times \mathbf{B}$ is at a minimum because the velocity $\mathbf{v}_s$ is at its minimum. Immediately afterwards, in the period 12:34:30–12:36:30, Fig. 3 shows that there is a moderate flux of ions travelling northwards with $\theta < 45^\circ$. Spectrograms for these ions have been plotted separately in Fig. 1, which is otherwise identical to Fig. 3. These ions show a steady rise in energy from 10 eV to 3 keV between 12:34:30 and 12:36:30 and, as well as flowing northwards, have a strong antisunward motion ($\phi = 180^\circ$). This population is separated in velocity space from the solar wind, which is still present near the ecliptic and neither starts this interval at such a low energy nor ends it at such a high one. Since the northward-moving population is not solar-wind plasma, it is, presumably, barium ions, accelerated by the electric field. Its behaviour is strikingly similar to that of lithium ions seen by the UKS in an earlier chemical release⁶. The electric field calculated from $-\mathbf{v}_s \times \mathbf{B}$, was northward throughout this time, which agrees with the direction of flow of these ions. If the electric field is assumed to be roughly constant, then the increasing energy of the barium ions is evidence that the source region is moving away from the satellite. If the average electric field is approximately equal to $-\mathbf{v}_s \times \mathbf{B}$ (8.2 mV m^{-1}) during this period, then the cloud's expansion velocity away from the satellite is 2.5 km s^{-1} along the electric field direction.

Conclusions

The measurements made by the UKS on the fringes of the diamagnetic cavity show that the solar wind was decelerated and deflected antiparallel to the electric field by the increased magnetic field outside the cavity. Apart from a short episode near the time at which the measured solar-wind speed was at a minimum, the barium ions were observed to be accelerated northwards by the electric field. The momentum of the barium ions must balance that of the southward deflection of the solar-wind ions.

The southward deflection of the solar wind brought the solar-wind ions through the diamagnetic cavity because the magnetic field strengths were not sufficient to deflect them away from the cavity⁵. Some of their momentum was transferred to the barium

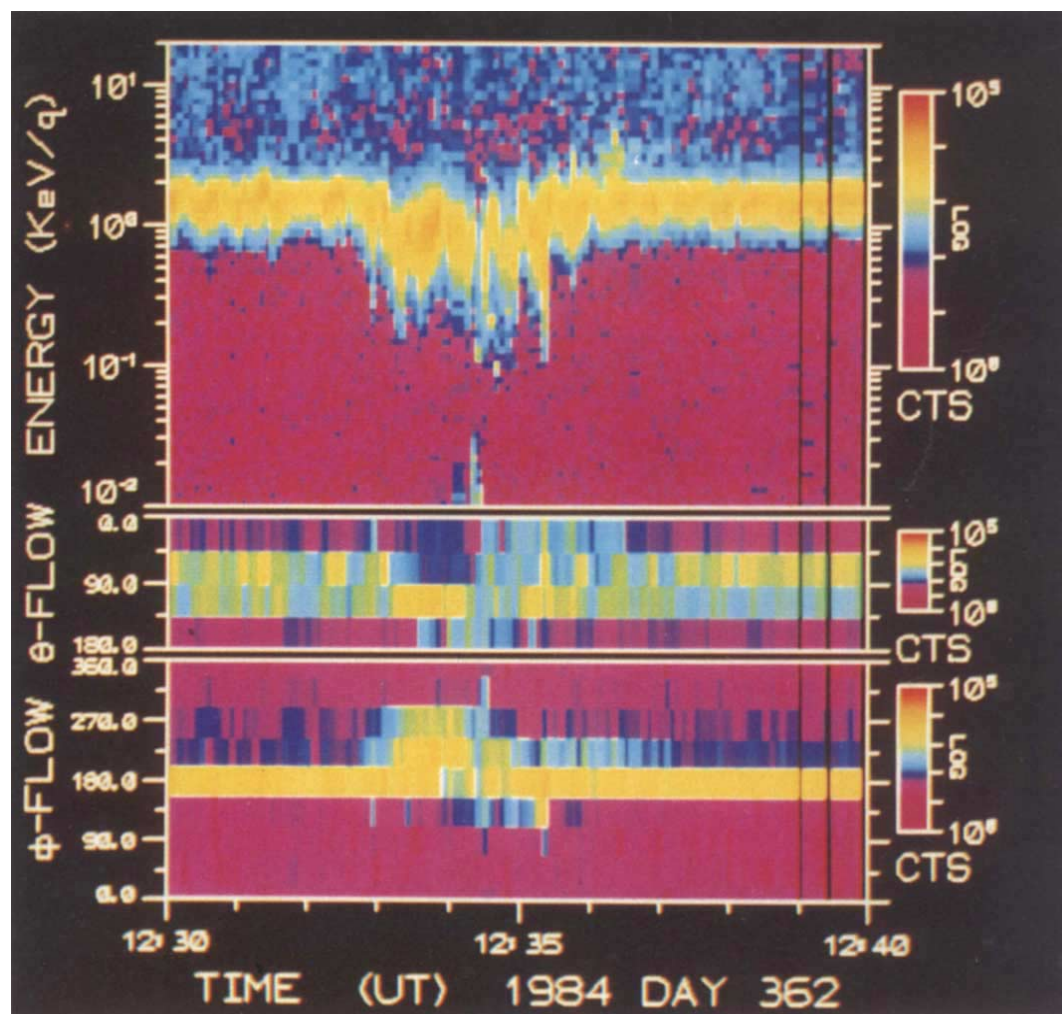


Fig. 3 Ion spectrograms of energy, 10 eV–20 keV (top), polar flow angle θ (middle) and azimuthal flow angle ϕ (bottom), against Universal Time. North is $\theta = 0$; sunward is $\phi = 0$. The colour scale represents normalized count rate, from 10^0 to 10^5 .

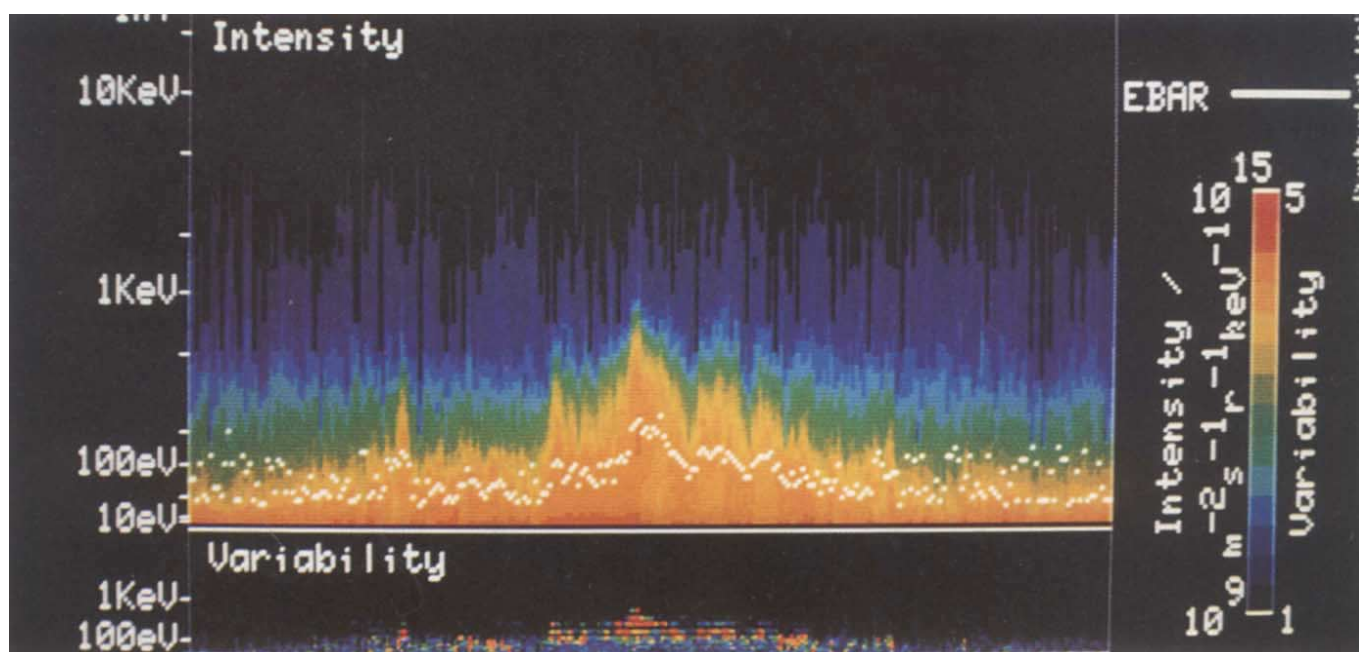


Fig. 4 Electron spectrograms. The top panel shows energy (12 eV–18 keV) against time with intensity given by the colour scale. In the bottom panel the colour scale represents variability (variance/expected variance), over the same time and energy ranges.

in the cavity itself, thereby leading to the observed southward motion of the cloud⁷.

Because the momentum loss of the light solar-wind ions in their initial deflection is balanced by the increased momentum of the very much heavier barium ions being accelerated northwards, a surplus of energy results. This energy is taken up by

the electrons, which do not play a significant part in the momentum equation. The electrons therefore play a similar role to the neutrinos in β -decay by acting as a repository for the energy remaining after momentum has been conserved. The spectral variations of the electrons indicate that the heating process is a resonant wave-particle interaction.

Received 13 December 1985; accepted 14 March 1986.

1. Lühr, H. *et al.* *Nature* **320**, 708-711 (1986).
2. Carlsen, J. L. *Planet. Space Sci.* **23**, 53-60 (1975).
3. Coates, A. J. *et al.* *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 287-292 (1985).
4. Shah, H. M., Hall, D. S. & Chaloner, C. P. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 293-300 (1985).

5. Haerendel, G., Paschmann, G., Baumjohann, W. & Carlson, C. W. *Nature* **320**, 720-723 (1986).
6. Johnstone, A. D., Coates, A. J., Smith, M. F. & Rodgers, D. J. *Trans. Chapman Conf. Ion Acceleration* (American Geophysical Union, in the press).
7. Rees, D., Hallinan, T. J., Stenbaek Nielsen, H. C., Mendiola, M. & Baumgardner, J. *Nature*, **320**, 704-708 (1986).
8. Bryant, D. A., Hall, D. S. & Chaloner, C. P. *Eur. Space Ag. spec. Publ.* **SP-229**, 229-304 (1985).

Plasma waves and wave-particle interactions seen at the UKS spacecraft during the AMPTE artificial comet experiment

L. J. C. Woolliscroft*, M. P. Gough†, P. J. Christiansen†, A. G. Darbyshire*, H. G. F. Gough‡, D. S. Hall§, D. Jones‡, S. R. Jones*|| & A. J. Norris†

* Department of Physics, The University of Sheffield, Sheffield, S3 7RH, UK

† School of Mathematical and Physical Sciences, The University of Sussex, Falmer, Brighton BN1 9QH, UK

‡ Space Plasma Physics Group, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

§ Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

Wave and particle correlator experiments on board the UKS revealed a wide range of plasma wave and wave-particle interactions associated with the 27 December 1984 artificial comet release just outside the main diamagnetic cavity. Shortly after the release strong bursts of plasma waves were recorded, and high-frequency electron oscillations in an enhanced-density plasma were observed later, along with evidence for interaction between a beam of electrons and the local plasma. At lower frequencies, extremely strong wave emissions were associated with what appeared to be the interface between the cloud and solar-wind plasmas.

IN this paper we discuss the plasma waves and wave-like modulations in the plasma electrons detected at the United Kingdom Satellite (UKS) during the first barium release of the AMPTE (Active Magnetosphere Particle Tracer) mission. The barium release, which took place as part of the active phase of the mission¹, occurred in the solar wind at 12:32:01 Universal Time (UT) on 27 December 1984. Details of the release physics are discussed by Haerendel^{2,3} and Dum⁴. The AMPTE UKS spacecraft was ~170 km away from the Ion Release Module (IRM) and there was no evidence (see ref. 5 for details) that the UKS entered the diamagnetic cavity. In order to interpret the observations it is necessary to understand the operation of the instruments, in particular the plasma wave and particle correlator experiments.

Instrumentation

The UKS wave experiment has been described by Darbyshire *et al.*⁶ and is capable of many modes of operation. The mode used during this release provided measurements of the electric component of the wavefield using signals from the dipole antenna. This dipole has a tip-to-tip separation of 7 m between carbon-coated spheres which contain preamplifiers and are connected to a stepped frequency analyser, four low- and medium-frequency bandpass filters, and an autocorrelator. This short dipole length was imposed by vehicle dynamics. The stepped frequency analyser covered the spectrum between 0.4 and 131 kHz, stepping downwards in frequency with 256 steps of

515-Hz bandwidth, the full sweep time being 16 s. The four bandpass filters were used for the processing of signals above the frequency range of the stepped frequency analyser. The centre frequencies of the filters were 222, 636, 1,140 and 2,130 kHz with ~5% bandwidth.

Figure 1 shows the output from the stepped frequency analyser for the whole period of the release; the wave amplitude is plotted as grey level against frequency and time. The constant-grey-level horizontal lines are residual interference lines, as is a broad band of noise at ~14 kHz. Also shown is the output of the 222-kHz filter for the short period 12:34:14-12:34:20 UT.

Fourier analysis of the output of the wave autocorrelator produces a 64-point spectrum, containing the frequency range 0.11-3.9 kHz, 7 times per second. The zero lag value of the autocorrelation function is

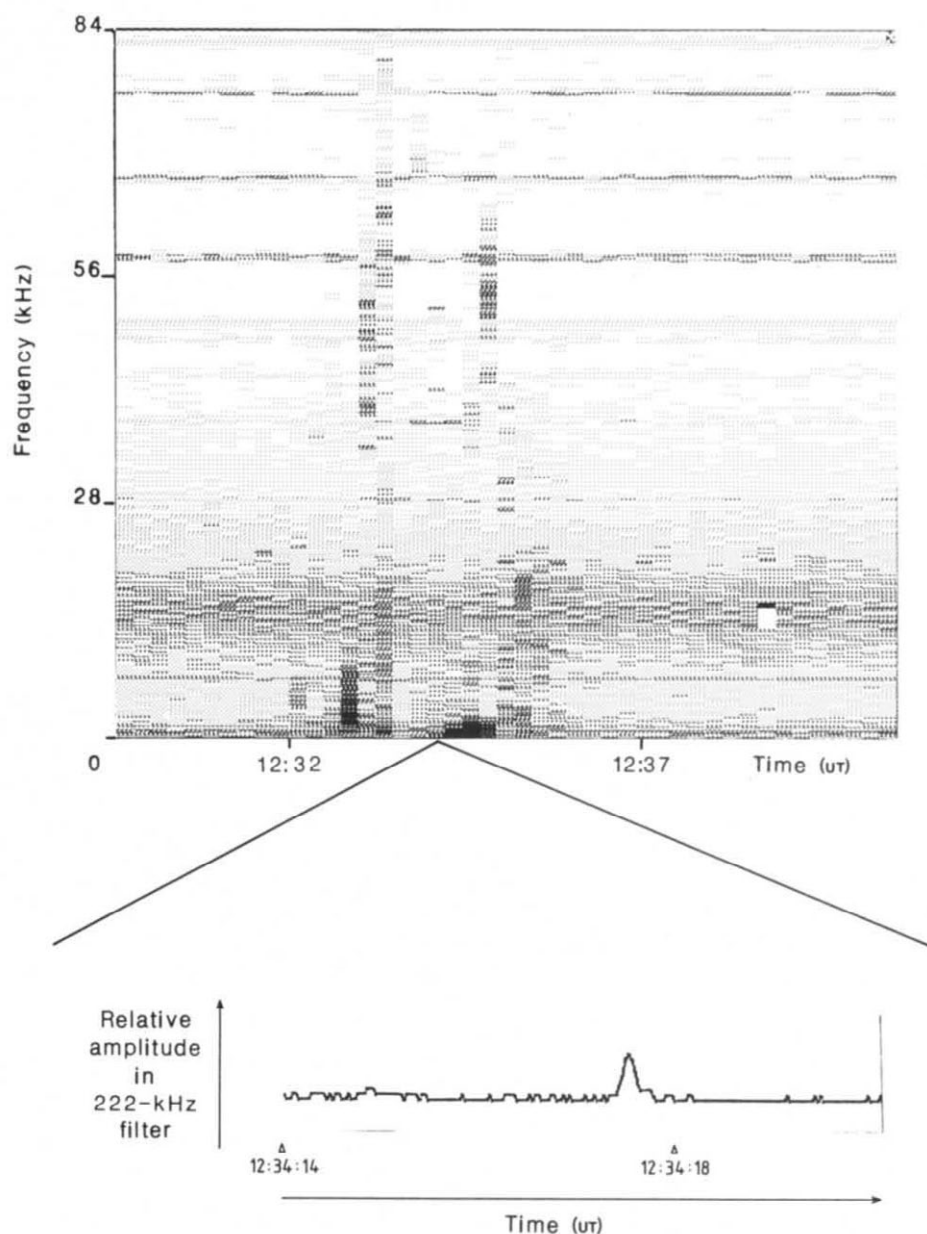
$$R = \frac{1}{N} \sum_{k=1}^N x(k\Delta t) x(k\Delta t)$$

where N is the number of samples of the waveform, and $x(k\Delta t)$ is the value of the k th sample, and thus is a measure of the total power within the 0.11-3.9 kHz band. Figure 2 is derived from these values, and shows the variation in power spectral density for the period 12:31 to 12:38 UT.

The particle correlator is a novel instrument which uses on-board microprocessors to perform analyses of data, thus extracting information which would be lost by conventional processing and telemetry. The two particle correlators take the output from the preamplifiers of the ion or electron particle experiment^{7,8}, and compute an energy-time matrix by an autocorrelation (acf) technique; that is, for each energy in the particle energy spec-

|| Present address: Satellites International Ltd, Bridge Road, Chertsey KT16 8LJ, UK.

Fig. 1 Grey-scale plot, against time, of the intensity of wave activity at different frequencies, measured by the UKS wave experiment stepped frequency analyser. Certain horizontal lines, such as those at 60 and 76 kHz, are the product of residual interference, as is a band of noise a few kHz wide at ~14 kHz. The barium release took place at 12:32:01 UT. The inset shows the output of the 222-kHz filter from 12:34:14 to 12:34:20 UT.



trum, any periodicities in the particle flux (in the kHz range) are recorded. The use of the autocorrelation technique allows integration until a significant number of counts have been collected. This is normally over several rotations of the spacecraft (the UKS spin rate was 12 r.p.m.), and during the artificial comet release a full energy set of correlation functions for each of the two correlators was telemetered in 16 s. Both correlators were configured to observe modulations of electrons arriving in the UKS spin plane, which was close to the plane of the ecliptic. The frequency ranges covered by the particle correlators were 0–60 kHz with 900-Hz resolution, and 0–2.2 kHz with 34-Hz resolution.

Subsequent fast Fourier transformation by ground-based computer provided detected modulations in an energy-frequency matrix. Each 16-s energy-frequency matrix was scanned for modulations of high (>99.9%) significance, which were then plotted, as in Fig. 3, as a function of energy and frequency versus time. The particle correlator experiment has been described more fully by Gough⁹ and is a sensitive way of detecting wave-particle interactions when the wave spectrum is more broad-band and of lower amplitude than would be obvious

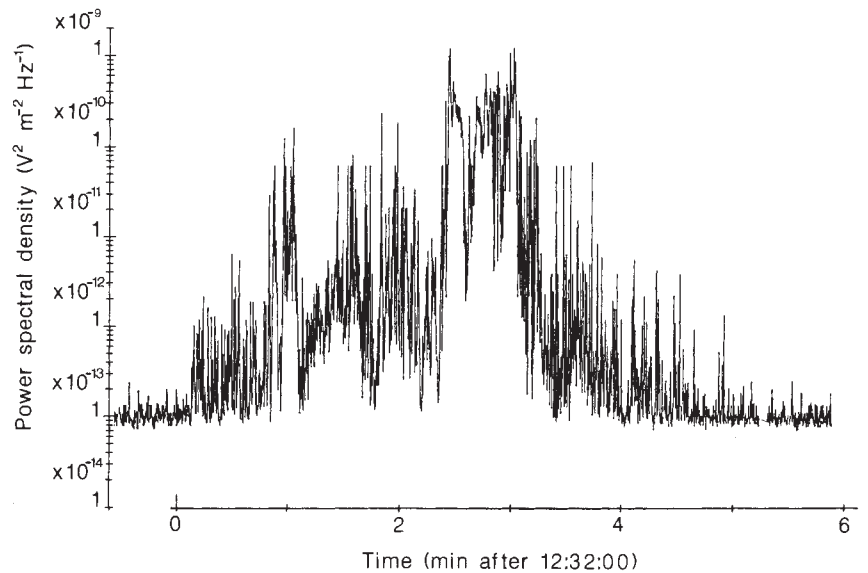
from use of a conventional combination of separate wave and particle experiments.

Results

There follows a description of the observed wave and particle modulation phenomena in a time-ordered sequence, along with an initial interpretation. Before the barium release, the wave experiment detected no plasma waves of natural origin because the short dipole length, compared with the Debye length, imposes a relative insensitivity and poor coupling of the antenna to the plasma when the electron number density is low. This depends on the ambient plasma conditions and is not always the case in the solar wind.

Similarly, the high-frequency particle correlator (0–60 kHz) did not detect significant modulations before the release (Fig. 3a), possibly because the solar-wind electron count rate is low. However, the 0–2 kHz particle correlator did detect modulations before 12:33 and after 12:36 UT (Fig. 3b). At the time of the release, the UKS was located upstream of the Earth's bow shock, in the region of the proton foreshock. These low-frequency modulations probably resulted from wave-particle interactions

Fig. 2 Power spectral density (assuming a flat spectrum) received in the 0.11–3.9-kHz band, calculated from the zero lag of the UKS wave experiment autocorrelator. Because the boom length of the UKS was rather short, the antenna may not be well coupled to the plasma; these figures therefore represent lower limits to the wave activity.



associated with the back-streaming electrons and protons which constitute the foreshock region. Such modulations are frequently observed in this region.

About 10 s after the release, a short burst of wave noise was detected by the stepped frequency analyser, and there was also a step in the signal level in the wave autocorrelator (Figs 1 and 2). This signal may have been due to the naturally occurring ion acoustic waves which populate the proton foreshock, although these waves were not seen before or after the release period. Freely expanding plasma from the release cloud was constrained by polarization effects to velocities of the order of the ion sound speed, a few km s^{-1} at most; therefore, 'plasma' contact with the UKS should have been established after a longer interval (of the order of a minute). The cloud may have communicated more rapidly through effects associated with field-line distortion and low-frequency Alfvén wave generation; however, there was little evidence of this at the UKS at this time.

At about 12:33:00 UT a strong burst of waves was detected at the UKS, again with frequencies below ~ 10 kHz, and with peak amplitude at ~ 3 kHz (Fig. 1). This appears to have been associated with the first detection of heated (~ 100 eV) electrons at UKS. Figure 2 shows that there were three peaks within 20 s; each of these peaks coincided with an enhancement in the count rate of electrons detected at the UKS. At the same time, the magnetic field at the UKS was observed to move significantly away from the solar-wind magnitude and direction and the solar-wind ions were observed to be retarded. We suggest that these electrons may have been heated in a turbulent region close to the solar wind/cloud boundary and that the UKS was magnetically connected to that region by the now-distorted field lines.

The first direct evidence from the wave experiment for enhancement in electron number density occurs in the stepped frequency analyser sweep at $\sim 12:33:24$ UT, when electron oscillations at 8–50 kHz were observed. This signal was extremely variable in frequency on short timescales, differing considerably between successive sweeps, and within a single sweep (Fig. 4, where data at frequencies of known interference lines are suppressed). If we make the preliminary assumptions that these were close to the electron plasma frequency, f_{pe} , and that these waves were driven by free energy of beam-like features in the electron distributions, then they corresponded to electron number density variations in the range $20\text{--}90\text{ cm}^{-3}$. After $\sim 12:33:50$, electron oscillations were mostly outside the range of the stepped frequency analyser until the period of decreasing density was reached at $\sim 12:35:00$ UT. The assumption that the frequency was close to f_{pe} may not be correct in certain rather

specialized circumstances¹⁰. Detailed particle correlator analysis is concentrated in a short period of 36 s, 12:33:34–12:34:10 UT, when the magnetic field was reasonably constant for longer than the integration period of the instrument. High-frequency modulations in the electron flux recorded in this period are plotted in Fig. 5 as a function of wave modulation frequency, f , against reciprocal wavelength, $k(=\lambda^{-1})$, on the assumption that the

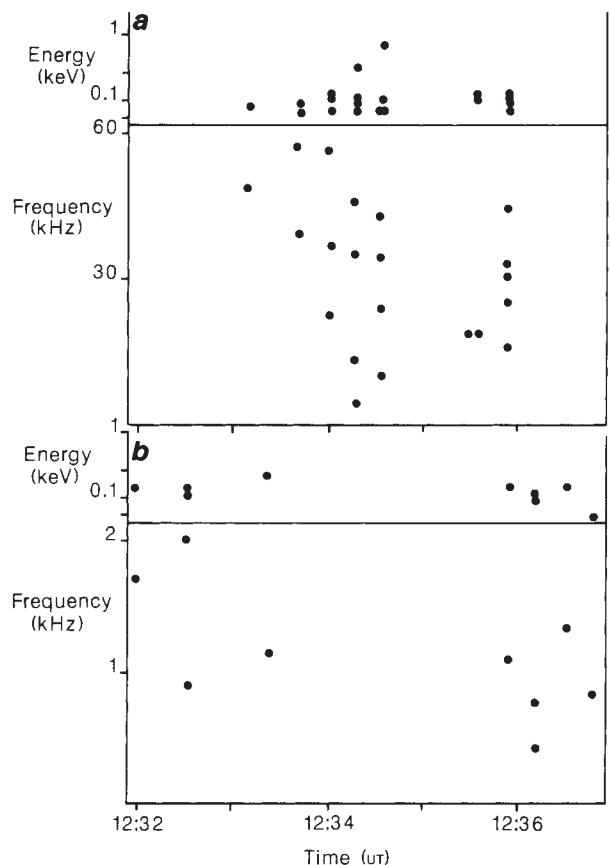
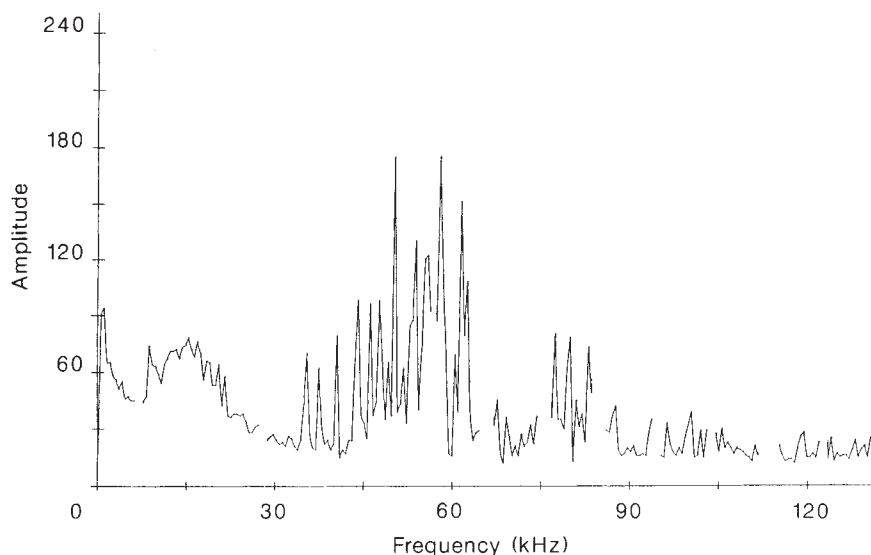


Fig. 3 Dots represent the presence of $>99.9\%$ -significant autocorrelations from the two particle correlators studying electrons arriving in the spin plane of the UKS. *a*, Data from the 0–60-kHz correlator; modulations occurred between 12:33:00 and 12:36:00 UT over a wide range of frequencies and at energies below 1 keV. *b*, Data from the 0–2.2 kHz correlator, showing modulations in electrons with energies of ~ 100 eV.

Fig. 4 Data from a single sweep of the wave experiment stepped frequency analyser for the period around 12:34:38 UT. Data at frequencies of known interference lines are suppressed. Note the highly structured nature of the strong electron oscillation noise between 40 and 80 kHz.



waves responsible for the electron modulations were in velocity resonance with the electrons ($f = v_e k$). All of the modulations lie on a curve, and the plot can be interpreted as an ω - k wave dispersion measurement. The measured points lie between the electron gyrofrequency, $f_g \approx 1.8$ kHz (from the UKS magnetometer) and the rapidly varying electron plasma frequency ($f_{pe} \approx 40$ –222 kHz). The appearance of highly variable enhanced electron fluxes in the 50–200 eV energy range during this period suggests a preliminary model in terms of a 'beam-plasma' interaction, the 'beam' of enhanced electrons being assumed to have originated in some low-frequency turbulent region associated with the expanding barium cloud. Although a fuller interpretation can be made only when high-time-resolution electron distribution function data are analysed, the preliminary data already reveal the presence of significant phase-space structuring in the relevant electron velocity range. The experimentally measured slope of the wave dispersion curve (Fig. 5) gives the electron beam energy in the range 50–150 eV.

At 12:34:18 UT electron oscillations were detected in the 222-kHz filter (Fig. 1), implying a plasma density of 600 cm^{-3} , which appears to be the peak value seen at UKS. This figure is higher than the number density derived from either the ion or electron experiments, and suggests the presence of substantial numbers of particles with energy below the lower limits of the plasma instruments. These oscillations coincided with the brief detection of low-energy (presumably barium) ions at UKS, and the maximum electron intensities. At the same time (within a few seconds), the total magnetic field increased by 20 nT to ~ 100 nT (see ref. 5), followed by a rapid drop to 30 nT. The plasma kinetic pressure, nkT (taking $n = 200 \text{ cm}^{-3}$, $T = 1$ eV, and where k is the Boltzmann constant), was by now comparable with the magnetic pressure (taking $B = 100$ nT), so that finite diamagnetic effects associated with the high-density front were not surprising. The decrease in field strength was precipitous (of the order of 10 nT s^{-1}), implying a rather large diamagnetic current density, $J \approx \Delta B / \mu_0 L_{\text{scale}}$ (where ΔB is the change in magnetic field and where μ_0 is the permeability of free space). The scale length, L_{scale} , is not known with any precision; nevertheless, it appears that the electron drift velocity needed to carry this current could have exceeded the threshold for ion acoustic instability for $L_{\text{scale}} < \text{several tens of kilometres}$ (see ref. 4).

After the density peak, electron oscillations were detected by the stepped frequency analyser at various frequencies during its next few sweeps, as can be seen in Fig. 1, where the strong feature at 50 kHz at $\sim 12:35:16$ straddled lower-frequency signals at earlier and later times. The electron number density was clearly highly variable. Indeed, the period 12:34:18–12:35:40 was characterized by strong fluctuations in the direction and magnitude of the magnetic field, electron intensities and energy

and direction of arrival of detected solar-wind ions.

During this period the most intense low-frequency (< 4 kHz) wave powers were observed (Fig. 2), corresponding to at least 2 mV m^{-1} broad-band. These can also be seen in the bottom few stepped frequency analyser steps. The temporal evolution showed several major intensity variations and the frequency evolution showed a generally rising frequency for this noise band. The precise relation between these bursts and variations on similar timescales in all other plasma parameters awaits further detailed analysis. At present the typical wavelengths are unknown, but if they were as low as 100 m then the parameter $e\phi/kT$ (where e is the electronic charge and ϕ is the amplitude of the electric component of the wave), which compares wave and kinetic potentials, may have been very large (≥ 0.1), implying strong local wave-particle interactions.

Intense noise was also recorded at this time at the IRM, and the general similarity in the magnetic field data at the two spacecraft show that they were probably in a similar region during this period. Gurnett *et al.*¹¹ discuss the generation of this noise in terms of an electrostatic shock and an ion beam-plasma interaction between the released barium ions and the solar-wind protons.

The fact that the 0–2-kHz particle correlator did not resolve quasi-monochromatic frequency components during this period is not surprising, in view of the turbulent nature of the wave spectrum.

Conclusions

The wave and particle correlator results have been analysed with only moderate time-resolution (a few seconds). However, they already reveal the existence of strong localized wave-

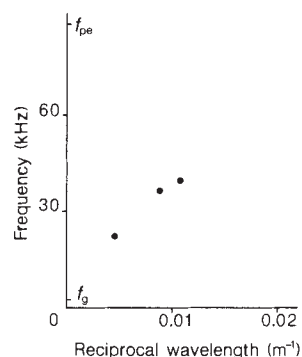


Fig. 5 Data for the period 12:33:34–12:34:10 UT from the 0–60 kHz particle correlator, displayed as a frequency versus reciprocal wavelength (ω - k) diagram, assuming that the modulations are in velocity resonance with the electrons.

particle interactions, particularly during the period in which a relatively dense and possibly somewhat fragmented portion of the original cloud expanded past the UKS. More precise identification of these local instabilities awaits comparison of data with that of other experiments at the highest possible time resolution. In addition to those instabilities driven by relative drift between electrons and ions, other waves driven by relative

motion between the cloud and solar-wind ions are possible¹¹.

The data discussed here were obtained from instruments which were largely produced by teams at the Universities of Sussex and Sheffield. Important contributions to the hardware and software were made by many colleagues, including A. G. Whitehurst, G. A. Wilson and J. A. Thompson at Sheffield, E. J. Gershuny at Sussex and I. Ferebee at Surrey University.

Received 10 December 1985; accepted 7 March 1986.

1. Krimigis, S. M., Haerendel, G., McEntire, R. W., Paschmann, G. & Bryant, D. A. *EOS* **63**, 843-850 (1982).
2. Haerendel, G. *Eur. Space Ag. spec. Publ. No. SP-195*, 337-340 (1983).
3. Haerendel, G., Paschmann, G., Baumjohann, W. & Carlson, C. W. *Nature* **320**, 720-723 (1986).
4. Dum, C. T. Preprint, Max-Planck-Institut für Physik und Astrophysik, Institut für Extraterrestrische Physik (March 1984).

5. Lühr, H. *et al. Nature* **320**, 708-711 (1986).
6. Darbyshire, A. G. *et al. IEEE Trans. Geosci. Remote Sensing* **GE-23**, 311-314 (1985).
7. Coates, A. J. *et al. IEEE Trans. Geosci. Remote Sensing* **GE-23**, 287-292 (1985).
8. Shah, H. M., Hall, D. S. & Chaloner, C. P. *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 293-300 (1985).
9. Gough, M. P. *IEEE Trans. Geoscience and Remote Sensing* **GE-23**, 305-310 (1985).
10. Mikhailovskii, B. *Theory of plasma instabilities* Vol. 1 (Consultants Bureau, New York, 1974).
11. Gurnett, D. A. *et al. Geophys. Res. Lett.* **12**, 851-854 (1985).

Dynamics of the AMPTE artificial comet

G. Haerendel*, G. Paschmann*, W. Baumjohann* & C. W. Carlson†

* Max-Planck-Institut für Physik und Astrophysik, Institut für extraterrestrische Physik, 8046 Garching, FRG

† Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

Sweeping the plasma before it, the interplanetary magnetic field quickly penetrated into an initially diamagnetic barium plasma cloud. The field was strongly compressed and extended in the flow direction. Ions accelerated by electric polarization fields formed a visible tail at the rear of the cloud; their recoil balancing the magnetic stresses. The dominant lateral motion of the head of the artificial comet is attributed to a recoil of ions extracted by the interplanetary electric field.

THE AMPTE (Active Magnetospheric Particle Tracer Explorers) mission artificial comet experiments on 27 December 1984 and 18 July 1985 created an excellent opportunity to study in detail a class of plasma interactions which occurs frequently in the universe and finds its most striking visible expression in the formation of the plasma tails of comets. (Unmagnetized planets such as Venus also seem to have similar tails.) The general situation is the interaction of a highly supersonic, dilute, magnetized plasma with a dense, stagnant, unmagnetized cloud. Molecular clouds in regions of star formation exposed to the expanding gases of a nearby hot star belong to the same category. The AMPTE experiments can contribute to the understanding of some of the aspects of this interaction; in particular, those concerning the magnetic field and momentum transfer. Naturally, our ability to simulate a cosmic phenomenon is rather limited. The release of 2 kg of Ba vapour, which becomes quickly ionized by solar ultraviolet radiation, creates a comet head whose size is limited by the most fundamental length scales of the

interaction of two dilute, collision-free plasmas; namely, ion inertial length and ion gyroradius. As a consequence the results derived from these experiments bear primarily on the microphysics of the plasma-solar wind interaction. This, on the other hand, is the aspect that is least accessible by remote astrophysical observations.

In this article we aim to give a unifying interpretation of the dynamical behaviour of plasma and fields during the artificial comet experiment. In addition to the data presented in the five companion papers,¹⁻⁵ we include *in situ* plasma measurements obtained with the Ion Release Module (IRM). We discuss the phenomena in their order of occurrence: (1) diamagnetic cavity, (2) return and compression of the magnetic field and (3) lateral displacement of the comet and formation of the tail. Phase 1 lasted ~60 s, and phase 2 occupied the following ~40 s. Phase 3 overlapped with phase 2, and continued for most of the comet's lifetime.

The MPE/UCB 3-dimensional (3D) plasma instrument,

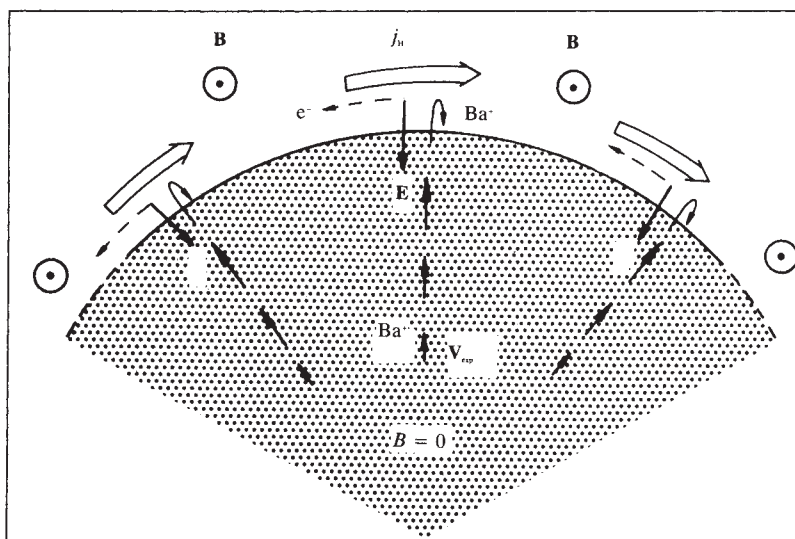


Fig. 1 Sector of the expanding barium plasma, creating a magnetic cavity. The shielding current, j_H , is carried by electrons in the boundary layer performing $\mathbf{E} \times \mathbf{B}$ drifts in response to the electric field, which retards the ions and maintains charge neutrality.

which made the measurements presented below, consists of two nearly identical electrostatic analysers, which during one spin period (~ 4.5 s) measure the complete velocity distribution of electrons and ions from 15 eV to 30 keV and from 20 eV to 40 keV, respectively⁶.

Magnetic cavity

The magnetometer data clearly show the creation of a magnetic cavity lasting for >1 min (ref. 3). This was precisely what was expected^{7,8} on the basis of earlier experiments⁹ and theory¹⁰. The cause of cavity formation is the high initial pressure of the expanding plasma cloud. Although the conversion of the neutral barium gas to a plasma proceeds with a time constant of ~ 30 s, the plasma density (and pressure) decays monotonically after an initial increase lasting for <1 s (ref. 5). Ions and electrons contribute almost equally to the total pressure of the plasma cloud. Because of charge neutrality constraints, both constituents are coupled together by an electric polarization field. Initially, the electron gas tends to get ahead of the ions and accelerates them by means of an outward-pointing radial ambipolar electric field. In this way the ions increase their maximum expansion speed from 1.0 km s^{-1} , stemming from the parent neutral atoms, to $\sim 1.35 \text{ km s}^{-1}$. The electron gas in turn is subject to adiabatic cooling.

The electrons interact directly with the ambient magnetic field. When they achieve pressure balance with it, the ions overshoot and set up an inward-pointing radial electric field. This retards the ions and leads to an electron current by virtue of their $\mathbf{E} \times \mathbf{B}$ drift. This current shields the interior of the cloud. Figure 1 is a sketch of these relations between the ions and electrons.

In order to understand the magnetic cavity of the artificial comets quantitatively, we must remember that the external plasma flow is supersonic relative to magnetosonic waves. This implies not only that dynamical equilibrium is reached with the external magnetic field, but that the kinetic pressure of the solar-wind ions is coupled in. This is again achieved through the action of an electric polarization field. The polarization field retards the solar-wind ions and drives an outer electron (Hall) current which generates a compressed magnetic field⁷. All of these features are reflected in both the IRM and UKS (United Kingdom Subsatellite) data.

Return of the magnetic field

After 77 s, the magnetic field at the IRM returned to almost the undisturbed level within <1 s and then proceeded to rise more slowly to even higher values³. If we take 60 s and 80 km as the time and radius of maximum extension of the diamagnetic cavity², we conclude that the average speed of the field penetration into the unmagnetized barium plasma was $\sim 4.5 \text{ km s}^{-1}$. In this context it is of greatest significance that, exactly coincident with the return of the field at the IRM, a sudden rise in plasma density by a factor of 5 was observed by the wave instrument⁵. Obviously, the magnetic field front, as it penetrates into the plasma cloud, sweeps the ions and electrons before it. The same observation has been made in dynamic pinch experiments and has been described in terms of a 'snowplough' model¹¹.

The ions have gyroradii that are large compared with the relevant scale lengths. For instance, the timescales of the two steps of initial field return are 0.3 and 3 s, respectively³. With a 'snowplough' (field penetration) speed of 4.5 km s^{-1} , this translates into spatial scales of 1.4 and 14 km. The gyroradius of a Ba ion moving with snowplough velocity even in the maximum field observed (130 nT) is 50 km. Thus, the ion has no difficulty in entering the magnetic field region as it sweeps across the cloud. It is the electrons which present a problem. Before we turn to this subject we try to understand the speed of the snowplough.

The Maxwell stresses, both normal and tangential, of the field draped around the barium cloud constitute the driving force. The electron and ion pressures are small after maximum

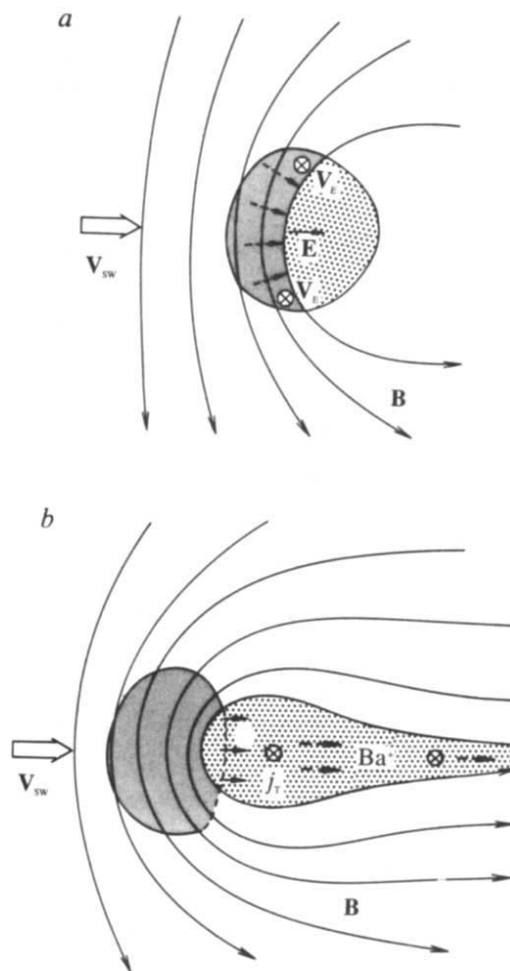


Fig. 2 *a*, Sketch of the 'snowplough' effect as the magnetic field penetrates into the initially unmagnetized plasma cloud. An electric field set up in the magnetic front drags the ions with it. Their inertia balances the magnetic stresses. The electrons experience an $\mathbf{E} \times \mathbf{B}$ drift— vE —which constitutes the shielding current. *b*, After magnetization of the whole cloud, the magnetic stresses act on the ions at the rear flank by means of an antisunward-pointing electric field. It injects the ions into the tail and drives a perpendicular electron (Hall) current ($j_{\perp}$) separating the antiparallel magnetic field in the tail lobes.

expansion. The only balancing force is the inertia of the ions as they are dragged with the magnetic front by means of an antisunward-pointing electric polarization field (Fig. 2*a*). This sweeping action causes the observed density rise. If we balance magnetic stresses and ion inertia, we can immediately derive an instantaneous propagation speed of the snowplough. It should be of order

$$v_{\text{SP}} = \frac{2B_c}{\sqrt{4\pi\rho_c}} \quad (1)$$

where the subscript 'c' refers to the compressed values. The factor of 2 occurs for the combined action of magnetic normal and shear stresses. With $B_c = 130 \text{ nT}$ (ref. 3) and $\rho_c = 1.2 \times 10^4 m_{\text{Ba}} \text{ cm}^{-3}$ (where m_{Ba} is the mass of the Ba ions; ref. 5), we derive $v_{\text{SP}} = 4.4 \text{ km s}^{-1}$, in excellent agreement with the estimate above.

The greatest surprise in the context of the observed field return was the absence of substantial wave noise⁵. We had anticipated an instability driven by the current along the boundary of the magnetic cavity, which would generate wave turbulence and anomalous electrical resistivity^{7,8}. This would allow a diffusive entry of the magnetic field into the cloud. However, merely in order to explain the first rise of B over a scale of $\sim 14 \text{ km}$ in

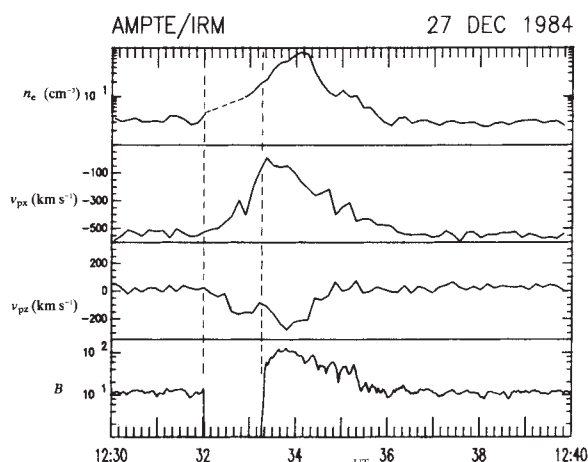


Fig. 3 Temporal development of hot electron (energy >15 eV) density (n_e) and proton bulk velocity components (v_{px} and v_{pz} in GSE coordinates; the x direction is toward the sun, z is toward the ecliptic north pole) during the first artificial comet release, together with magnetic field amplitude.

terms of magnetic diffusion, the diffusivity should have a magnitude of $\sim 6 \times 10^{11} \text{ cm}^2 \text{ s}^{-1}$. With the observed density and an average electron energy of ~ 1 eV this leads to expected electric field amplitudes of $\sim 10^3 \text{ mV m}^{-1}$, probably concentrated in the neighbourhood of the ion plasma frequency, 2 kHz. The wave measurements⁵ show only amplitudes well below 1 mV m^{-1} . There are at least seven orders of magnitude in wave power missing, if the penetration of B is to be attributed to a diffusive process.

How do the electrons manage to enter the magnetic field and neutralize the ions? The electric polarization field, which governs the dynamics of the essentially unmagnetized ions and which should be essentially normal to the magnetic front (see Fig. 2), forces the electrons to drift tangentially to the front. On the other hand, there cannot be any doubt that it is the cold (barium) photoelectrons which establish charge neutrality behind the front. The solar-wind electrons have much higher energy and would have been registered with the electron analyser, if they had been sucked in by parallel electric fields and piled up in the snowplough region. The ability of the electrons to penetrate far behind the magnetic front could be explained by the presence of small-scale structures caused by a generalized Rayleigh-Taylor instability. Such fine structure was resolved in a later experiment in the magnetic tail¹⁴. A small wave pulse near and above the electron gyro-frequency (~ 200 Hz) observed exactly upon entry into the magnetic field at 12:33:17 UT may be significant. It could be related to the first step of transferring electrons from the cavity into the magnetic field.

Compression of the magnetic field

The magnitude of the magnetic field compression can be readily understood in terms of the forces exerted by the solar wind. As the sketch of the field draping in Fig. 2 suggests immediately, magnetic normal and tangential stresses contribute about equally to the total force. This concept can be pursued in the framework of magnetohydrodynamic theory¹² which predicts a simple relation between maximum compression, κ , and alfvénic Mach number, $M_{A\perp}$, of the solar wind component perpendicular to $\mathbf{B}$:

$$\kappa = \frac{B_c}{B_0} = 2M_{A\perp} = 2 \frac{v_{sw\perp}}{v_A} \quad (2)$$

This is indeed consistent with the observations. On 27 December 1984, $M_{A\perp} \approx 6$ and $\kappa = 12$ was observed. On 18 July 1985,

$M_{A\perp} \approx 3.3$, but the magnetic field had an appreciable component aligned with the flow. Thus $M_{A\perp}$ was ~ 2.3 and the observed compression was 4.5.

The energy stored in the compressed and draped field is derived from the solar wind. This is clearly exhibited by the average solar wind parameters, a few of which are shown in Fig. 3. A reduction of solar wind velocity, $-v_{px}$, begins to develop soon after formation of the cloud. The IMR spacecraft is still well inside the cavity and sees the result of a solar-wind ion retardation occurring at the front. At about 12:33 UT the retardation of the ions increases dramatically. Just when the magnetic snowplough passes the IRM, $v_{px} \approx 0$. At the same time, a negative z component has developed. This is a consequence of the enhanced Lorentz force in the compressed field. The gyroradius of a solar-wind proton in a 100-nT field would be ~ 50 km. Because of the retardation of v_{px} , the actual value is even smaller.

The hot electrons (energy >15 eV) are seen to increase in density up to about 50 cm^{-3} , whereas the proton density (not shown) drops. It must be remembered that the hot electrons constitute only a minority in a cold-plasma environment ($\sim 10^4 \text{ cm}^{-3}$ from 12:33:20 to 12:33:50 UT). This increase of electron density was also seen at the UKS in the wings of the distorted field and can be interpreted as a combination of adiabatic compression of the solar-wind electrons and heating of photoelectrons by resonant wave-particle interactions⁴.

Quantitative evaluation of these observations shows that the magnetic compression results only partially ($<50\%$) from the local retardation of the average solar-wind ion. A large fraction of the momentum is transferred to the comet head from the Alfvén wings.

Momentum transfer

Having measured the compression and draping of the magnetic field inside and around the barium cloud and being able to account for this in terms of momentum and energy extracted from the solar wind flow, we should be able to estimate the total acceleration of the cloud under the action of this force. A simple way of describing the momentum transfer is to formulate the characteristic time τ_{acc} , which is taken to accelerate the cloud to solar wind speed⁸:

$$\frac{dv_c}{dt} = \frac{v_{sw} - v_c}{\tau_{acc}} \quad (3)$$

It can be shown¹² that

$$\tau_{acc} = \frac{\rho_c l_c}{4\kappa \rho_a v_{Aa}} \quad (4)$$

where ρ_c and ρ_a are the densities inside and outside the cloud, l_c its diameter, v_{Aa} the ambient Alfvén velocity and κ the magnetic compression ratio. Apart from a factor $(2\kappa)^{-1}$, this is the expression derived earlier for the subsonic case; that is, for $\kappa = 1$ (ref. 13).

Inserting the characteristic cloud and ambient parameters, we derive acceleration times of 1.6 and 1.3 h, respectively, for the two experiments. This means that the comet head should have been displaced downstream by 2,700 and 1,600 km, respectively, in the course of 4 min. However, the first artificial comet showed no discernible displacement in solar wind direction during this period¹; the same holds for the second experiment. Instead we observe an unexpected transverse acceleration. This observation constitutes perhaps the most important finding connected with these experiments.

The explanation is in fact quite simple. The absence of a downstream displacement and the presence of a lateral acceleration can both be traced to the same effect: ion extraction. The force exerted by the solar wind is not applied to the whole body of the cloud (after it has been fully pervaded by the magnetic field; that is, after ~ 100 s), but rather to a small fraction of the ions at the inner magnetic front, the region where the compressed

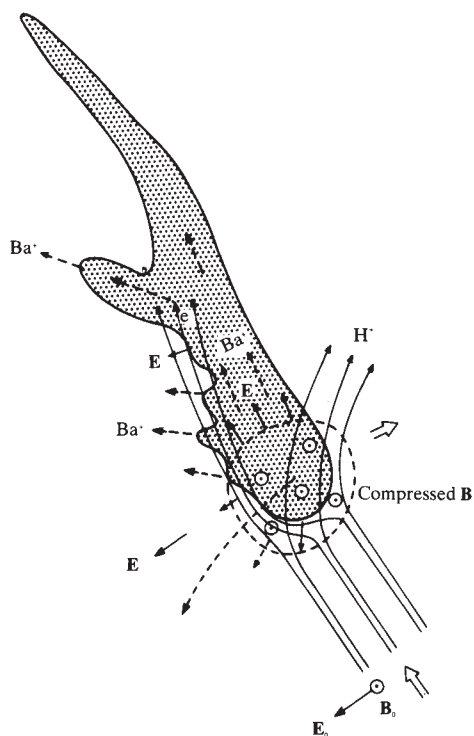


Fig. 4 Sketch of the interaction of the solar wind with the comet's head and tail. The density distribution is taken from Fig. 1 of ref. 1. Deflection of solar-wind ions and the recoil of the comet head balance the force extracting ions in the opposite direction. The strongly modified electric field guides the solar-wind electrons around the comet head. The recoil of the comet head causes an exit of the spacecraft (IRM) in the direction of E_0 .

field drops off on the rear side. Here, a strong electric polarization field exists which injects Ba ions at high speed into the tail (Fig. 2b). Apparently, there is no net force in the $-x$ direction inside the main cloud (comet head). The upstream-pointing magnetic pressure force is balanced by the tension of the draped field.

In the perpendicular direction the situation is different. Far from the barium cloud an electric field $E_0 = -(\mathbf{v}_{sw} \times \mathbf{B})/c$ exists (where c is the speed of light), pointing approximately in the $+z$ direction ($B_y > 0$). This field is highly perturbed inside the barium plasma. Nonetheless, it maintains on average its external orientation, albeit at lower magnitude. This field extracts ions both from the upper surface of the comet head and from the tail. The sharp lower edge of the tail of 27 December 1984, several upward-pointing features (protrusions), as well as a generally softer density decrease in the $+z$ direction all bear witness to this process. The ions extracted from the comet head in the $+z$ direction are actually observed by the 3D plasma instrument on IRM as a beam in the z direction; that is, along the electric field. As the spacecraft moves away from the barium cloud, the acceleration pathlength increases, and the mean ion energy increases accordingly.

It is the recoil of the tailward- and upward-extracted ions that prevents movement of the comet with the solar wind and accelerates both head and tail sideways, as long as the head density is sufficiently high. Balancing the total force on the comet by the

inertia of the ions injected into the tail and postulating that the tail density equals the ambient value, the injection speed into the tail is given by

$$v_{inj} = \frac{2}{\sqrt{\mu_{Ba}}} v_{sw} \quad (5)$$

where μ_{Ba} is the atomic mass of barium. This suggests $v_{inj} = 90$ and 45 km s^{-1} , respectively, in the two experiments and is consistent with motions derived from tracing features in the tail¹. The reason for assuming that the tail electron density is close to the ambient one is the need for charge neutrality of the injected ions. The neutralizing electrons at the same time constitute the sheet current, j_T , which separates the antiparallel fields in the tail lobes (Fig. 2b).

The lateral force caused by ion extraction by the (perturbed) solar-wind electric field from head and tail turns out to be $\sim 20\%$ of the force acting downstream. Recoil of the head and deflection of solar-wind ions in the $-z$ direction balance this force (Fig. 4). The loss of ions from the head to the tail by the discussed injection and the subsequent ion extraction explains the observed lifetime of 5 min.

Several observations communicated in the companion papers are closely related to the ion injection and extraction. First, the 3D plasma instrument sees the tail inject ion already in operation when the 'snowplough' passes the IRM. It appears as an antisunward-directed ion beam with energy $\leq 300 \text{ eV}$. However, here the process acts in the high density of the cloud's centre; the accelerating field is reduced by the high ion inertia. Only $\sim 20 \text{ s}$ later, when the front has reached the rear side of the cloud, the force acts on fewer particles and the accelerating voltage is correspondingly higher.

The recoil displaces the comet head in the $-z$ direction, so that the IRM spacecraft exits the cloud in the $+z$ direction; that is, in that of the ambient electric field (Fig. 4). This must be realized when interpreting the history of the particle and field parameters. It enabled us to detect the extracted Ba ions, the energy of which increased with growing separation from the comet head. Combination of the observed displacement with the energy increase of the extracted ions allows the derivation of the field increase on the upper side of the comet head. It is of the order of $10^{-2} \text{ V km}^{-2}$.

Conclusion

The interaction between the Ba plasma and the solar wind is dominated by the inertia of the injected ions in the presence of a magnetized electron fluid, and by the need to maintain charge neutrality. The Ba ions respond almost solely to the electric field, whereas the electrons perform an $\mathbf{E} \times \mathbf{B}$ drift and thus constitute a Hall current which largely determines the magnetic field distribution. The force exerted by the solar wind is thus communicated by magnetic stresses to the electrons and further by electric polarization fields to the ions. The response of the bulk of the ions (the visible head and tail) to the invisible extracted ions is achieved in the same way by coupling of electric and magnetic forces. It is fascinating to observe the collective expression of this interplay in a well-structured phenomenon resembling in many ways real comets and their plasma tails.

We thank the authors of the companion papers as well as the many individuals acknowledged in them for their effort in making the experiment a success. Thanks are also due to the engineers and technicians at MPE and UCB who contributed to the 3D plasma experiment on IRM.

Received 2 December 1985; accepted 3 March 1986.

- Valenzuela, A. *et al. Nature* **320**, 700-703 (1986).
- Rees, D., Hallinan, T. J., Stenbaek Nielsen, H. C., Mendiola, M. & Baumgardner, J. *Nature* **320**, 704-708 (1986).
- Lühr, H. *et al. Nature* **320**, 708-711 (1986).
- Rodgers, D. J. *et al. Nature* **320**, 712-716 (1986).
- Gurnett, D. A. *et al. Geophys. Res. Lett.* **12**, 851-864 (1985).
- Paschmann, G. *et al. IEE Trans. Geosci. Remote Sensing* **GE-23**, 262-266 (1985).

- Haerendel, G. *Eur. Space Ag. J.* **7**, 135-144 (1983).
- Haerendel, G. *Eur. Space Ag. spec. Publ. No. SP-195*, 337-340 (1983).
- Haerendel, G. *Space Res.* **13**, 601-617 (1973).
- Pilipp, W. *Planet. Space Sci.* **19**, 1095-1119 (1971).
- Krall, N. A. & Trivelpiece, A. W. *Principles of Plasma Physics*, 123-128 (McGraw-Hill, New York, 1973).
- Haerendel, G. *J. geophys. Res.* (submitted).
- Scholerr, M. *Planet. Space Sci.* **18**, 977-1004 (1970).
- Haerendel, G. *Phys. Bl.* **40**(5), 134-137 (1986).

How many more discoveries in the Universe?

Martin Harwit

Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853-6801, USA

Roger Hildebrand

Enrico Fermi Institute, Department of Physics, and Department of Astronomy and Astrophysics, The University of Chicago, Chicago, Illinois 60637, USA

It is generally assumed that the potential scope of any scientific discipline is quite incalculable. Yet the discovery of the Earth's major rivers, mountains and seas, an enterprise that may have seemed endless to Ptolemy's contemporaries, is now essentially complete. The same may soon be said about the discovery of new plants and animals on Earth¹. Harwit^{2,3} has put forward a procedure for estimating the number of observational phenomena remaining to be discovered in astronomy. Here we describe a statistical approach which supports the conclusion that the number of phenomena already discovered represents an appreciable fraction of those that can be found.

Harwit's phenomena are defined by a list of 43 archetypes (see Table 1), and he estimates that this list contains approximately one-third of all observational phenomena of comparable importance that will ultimately become recognized. Under extreme conditions matters might be worse, and we might thus far have uncovered only a tenth of all discoverable phenomena.

The procedure for making these estimates is based on a Poisson statistical approach that has long been in use by ecologists, linguists and statisticians⁴⁻⁶. Harwit argues that the Universe we inhabit transmits radiation only over a finite range of energies; moreover, the spectral, spatial and temporal resolving powers for electromagnetic radiation and other carriers of information are fundamentally limited by the Heisenberg uncertainty principle, the nature of the interstellar medium, the finite baselines across which telescopes may be arrayed and other practical factors. Barring the possibility of interstellar or intergalactic space travel, there is a limited range of events which astronomers can observe. The limits are not imposed by a lack of ingenuity, but rather by the structure of the Universe, the laws of physics, and the kind of galaxy we inhabit. Had the Solar System been placed in an elliptical galaxy, rather than in our spiral Milky Way, limitations imposed by the interstellar medium would be less restrictive, and the number of observable phenomena could be greater.

The Poisson statistical approach assumes that individual cosmic phenomena are observable through distinct sets of observing techniques. An entirely novel spectral range, or range of resolving powers, provides a fresh new look at the Universe and may permit either the discovery of a previously unrecognized phenomenon or the recognition of a previously known phenomenon by quite independent means.

Just as Poisson statistics permit the estimation of the size of a complete set of objects from a count of the number of triplicates, duplicates and singles in a random collection, so the number of rediscoveries in astronomy—that is, the independent recognition of phenomena through widely differing channels of information—provides a guide to the total set of observational phenomena awaiting discovery. But is the estimate obtained by Poisson statistics robust, or do other estimation techniques lead to different conclusions?

To test that question we have subjected the data assembled by Harwit to a procedure devised by Derenzo and Hildebrand⁷ for estimating scanning efficiencies in the detection of high-

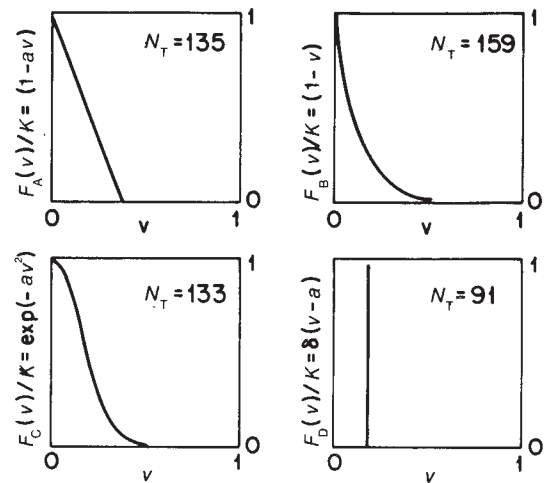


Fig. 1 Trial functions, $F_A - F_D$, as adjusted to fit the data for three observers. (See Table 2).

energy physics events. This procedure assumes that the visibility, v , the likelihood that a particular type of event (phenomenon) will be recognized by imperfect scanners (observers) using imperfect techniques, may range from $v = 0$ for invisible events to $v = 1$ for unmistakable events. If $F(v)$ is the number of events with visibilities in the range v to $v + dv$, then the total number of events should be

$$N_T = \int_0^1 F(v) dv \quad (1)$$

The function $F(v)$ is estimated from the results of several independent scans.

Consider a set of scanners viewing an identical data set. One scanner notes the existence of a number of events, A ; another recognizes B events, the third C , and so on. The order of the scans is of no significance. If there are four scanners, the average number of events detected between them can be labelled $M[1]$:

$$M[1] = [A + B + C + D]/4 \quad (2a)$$

If we now remove from the full data set those events seen by every one of the four scanners, then the average number found in the remaining set of less obvious events is

$$M[2] = [Ab + Ac + Ad + Ba + Bc + Bd + Ca + Cb + Cd + Da + Db + Dc]/12 \quad (2b)$$

where Ab stands for the number of events seen by scanner A but missed by scanner B , and so forth. If we next remove the events seen by all but one of the scanners, then the average number found in the remaining set of still less obvious events will be

$$M[3] = [Abc + Abd + Acd + Bac + Bad + Bcd + Cab + Cad + Cbd + Dab + Dac + Dbc]/12 \quad (2c)$$

and finally

$$M[4] = [Abcd + Bacd + Cabd + Dabc]/4 \quad (2d)$$

With just three scanners, the corresponding expressions would be

$$M[1] = [A + B + C]/3 \quad (3a)$$

$$M[2] = [Ab + Ac + Ba + Bc + Ca + Cb]/6 \quad (3b)$$

$$M[3] = [Abc + Bac + Cab]/3 \quad (3c)$$

The distribution of undetected events after one, two and n scans is $F[v][1-v]$, $F[v][1-v]^2$ and $F[v][1-v]^n$ respectively.

Table 1 Distribution of phenomena among observers

Archetypal phenomena	Four observers				Three observers			Archetypal phenomena	Four observers		Three observers	
	A	B	C	D	A	B	C					
1 Stars	×				×			26 Supernova remnants	×		×	26
2 Planets	×	×			×	×		27 Radio galaxies	×		×	27
3 Novae	×				×			28 Magnetic variables	×		×	28
4 Comets			×			×		29 Flare stars	×		×	29
5 Moons	×				×			30 Interstellar magnetic fields	×	×	×	30
6 Rings	×				×			31 X-ray stars		×	×	31
7 Galactic clusters	×				×			32 X-ray background		×	×	32
8 Clusters of galaxies	×				×			33 Quasars	×		×	33
9 Interplanetary matter	×		×	×	×	×	×	34 Microwave background	×		×	34
10 Asteroids			×				×	35 Masers	×		×	35
11 Multiple stars	×				×			36 Infrared stars		×	×	36
12 Variable stars with nebulosity	×				×			37 X-ray galaxies		×	×	37
13 Planetary nebulae	×	×			×	×		38 Pulsars	×		×	38
14 Globular clusters	×				×			39 γ -ray background		×	×	39
15 Ionized gas clouds		×	×		×	×		40 Infrared galaxies		×	×	40
16 Cold interstellar gas		×	×		×	×		41 Superluminal sources	×		×	41
17 Giants main sequence stars			×			×		42 γ -Ray bursts		×	×	42
18 Cosmic rays		×		×	×	×		43 Unidentified radio sources	×		×	43
19 Pulsating variables	×				×				$M[1] = 13$		$M[1] = 17.33$	
20 White dwarfs			×				×		$M[2] = 11.25$		$M[2] = 14$	
21 Galaxies	×				×				$M[3] = 10$		$M[3] = 11.33$	
22 Cosmic expansion			×			×			$M[4] = 8.5$			
23 Interstellar dust			×			×						
24 Novae/supernovae	×				×							
25 Galaxies with/without gas	×	×			×	×						

Four observers: A uses optical continuum methods; B uses radio techniques; C uses optical spectroscopy and astrometry; D uses all other techniques. Three observers: A uses optical continuum methods; B uses infrared radio, and electroscopie techniques; C uses optical spectroscopy, polarimetry, astrometry, X- and γ -ray detection and meteorite collection.

Table 2 Trial functions and corresponding estimates of N_T

Function	Parameters*		$N_T^\dagger$	
	Four observers	Three observers	Four observers	Three observers
$F_A = K(1 - av), v < a$ $0, v \geq a$	$K = 1,224, a = 4$	$K = 703, a = 2.6$	153	135
$F_B = K(1 - v)^a$	$K = 1,633, a = 10$	$K = 1,308, a = 7.2$	142	159
$F_C = K \exp(-av^2)$	$K = 1,082, a = 42$	$K = 690, a = 20$	148	133
$F_D = \delta(v - a)$	$K = 100, a = 0.13$	$K = 91, a = 0.19$	100	91
$F_E = K_1 \delta(v - a_1) + K_2(v - a_2)$	$K_1 = 400, a_1 = 0.0025$ $K_2 = 86, a_2 = 0.14$	— ‡	486	‡

* Values giving best fits to $M[i]$. See Table 3.

† See equation (1).

‡ No solution (four parameters, three equations).

Hence the number of new events found in the n th scan will be

$$M[n] = \int_0^1 cF[v] [1 - v]^{n-1} dv \quad (4)$$

The collection of values $M[1], M[2], \dots, M[n]$ can then be used to estimate the function $F[v]$. In general, the total number of events, N_T , derived from equation (1) will be approximately the same for any smooth trial function, $F(v)$, whose m parameters ($m \leq n$) are adjusted to give satisfactory fits to the values of $M[i]$.

As discussed by Derenzo and Hildebrand¹, the statistical techniques commonly used to estimate total populations from scanning samples correspond implicitly to the assumption the $F(V)$ is a single delta function. That assumption leads systematically to an underestimate of N_T ; it does not allow for the likelihood that different scanners will miss the same (relatively obscure) events. Harwit's simplest fit is an example².

If we now examine the data compiled by Harwit³, we note that his observing instruments can be lumped loosely into four different scanning categories without in any way changing the phenomena that are singly recognized, doubly recognized or triply recognized: with four scanners, one could assign all optical continuum methods, including the human eye, photographic techniques used on a telescope, and so forth, to scanner A; all possible radio techniques, continuum as well as spectral line observations, to scanner B; all optical spectroscopy and astrometry to scanner C; and all other techniques, including radar, electroscopes, X-ray, γ -ray and infrared detectors, and meteorite collection, to scanner D. Any lumping of observing instruments will give the same result provided only that one assigns to different scanners the instruments leading to rediscoveries. For example, no change in the values of M or N would result if scanners C and D exchanged radar and astrometric

Table 3 Fits to values of $M[i]$ from Table 1 with functions and parameters from Table 2

	Four observers				Three observers		
	$M[1]$	$M[2]$	$M[3]$	$M[4]$	$M[1]$	$M[2]$	$M[3]$
Observed*	13	11.25	10	8.5	17.33	14	11.33
F_A	13	11.7	10.1	8.8	17.33	14.0	11.5
F_B	13	11.1	9.6	8.4	17.33	13.9	11.4
F_C	13	11.2	9.8	8.6	17.33	13.9	11.4
F_D	13	11.3	9.8	8.6	17.3	14.0	11.3
F_E	13	11.35	9.9	8.7	—	—	—

* Based on data from ref. 3 (see table 1).

techniques or if all X-ray observations were made by C instead of D. Only minor changes would result if one used, say five scanners or three—we need at least three to take account of the triple coincidence identified by Harwit.

With the grouping as we have described it, the 43 discoveries are distributed among the four scanners as shown in the 'four observers' column of Table 1; a distribution for three scanners is shown in the last three columns. The corresponding values of $M[i]$ are shown at the bottom of Table 1.

In Table 2 we list five trial functions, the values of the parameters giving the best fits to the values of $M[i]$ from Table 1, and the corresponding estimates of the total number of phenomena. The functions are illustrated in Fig. 1, and the quality of the fits may be seen in Table 3. As anticipated the smooth functions give similar values and the single delta function, F_D , corresponding to Harwit's simplest case, gives a somewhat lower value. The function F_E , a fit to two delta functions, gives a value which is higher but not an order of magnitude higher; it is almost the same as the value derived in Harwit's most pessimistic analysis³. As our astrophysical understanding grows, improved lower limits on visibility can be derived, probably relieving the severity of this most pessimistic estimate and leading to lower estimates for N .

Throughout this discussion we have restricted ourselves to a consideration of the archetypal phenomena listed in Table 1. Doubts might therefore be expressed about the validity of that list as a representative sample of astronomical discoveries in general. However, several related criteria can lead to a list of almost identical length having virtually the same rediscovery statistics³. One alternative criterion that can be used to define cosmic phenomena similar to those of Table 1 specifies that classification as a distinct phenomenon can be defined strictly on the basis of observational criteria: Let a phenomenon exhibit itself primarily through radiation characterized by wavelength λ_1 , line width $\Delta\lambda_1$, modulation on a time scale Δt_1 , angular size in the sky $\Delta\theta_1$ and polarization characteristics P_1 . Then we define phenomenon 1 as distinct from all other known phenomena, j , with characteristics λ_j , $\Delta\lambda_j/\lambda_j$, Δt_j , $\Delta\theta_j$, p_j , if at least one of these five parameters differs by at least a factor of 10^3 from that characterizing any other phenomenon, j . Several ways of defining a phenomenon were suggested by Harwit³, their common ingredient being differentiation by a factor of 10^3 . All led to roughly similar Poisson statistical results, suggesting that the list of phenomena used (Table 1) is robust and can be considered representative of phenomena of a given degree of importance.

In this connection we should point out that a different archetypal list would have resulted if we had selected a different importance class. Phenomena differing by a factor of only 10^2 , or perhaps 10, in observational characteristics might be expected to be far more abundant and perhaps the ratio of present discoveries to ultimately expected discoveries would be quite different. That certainly would not be surprising.

One possible set of objections to our approach still needs to be considered. It might be argued that these conclusions are misleading because of an over-restrictive definition of a

phenomenon—that within astronomy there are many areas where a rather modest quantitative improvement in sensitivity can reveal qualitatively new effects. For instance, once it becomes possible to study galaxies at a distance such that the light travel time from them is a significant fraction of the age of the universe, the entire subject of evolutionary cosmology might be thought to open up. However, there is a fallacy in this argument. The next generation of orbiting infrared instruments will indeed enable us to investigate galaxies that are red-shifted by a factor of 10, and thus to study the evolution of galaxies. But these telescopes will be operating at wavelengths ten times longer than those now used for investigating galaxies, so that we will again be dealing with sources an order of magnitude different in appearance.

Thus, if we restrict ourselves to phenomena of the importance class represented by the archetypal list of Table 1, we find that the number of phenomena already discovered is an appreciable fraction of those we may ultimately expect to find, as astronomy matures and the taxonomy of major cosmic phenomena reaches completion.

Received 11 October 1985; accepted 20 February 1986.

1. Diamond, J. M. *Nature* **315**, 538–539 (1985).
2. Harwit, M. Q. *J. Soc.* **16**, 378–409 (1975).
3. Harwit, M. *Cosmic Discovery* (Massachusetts Institute of Technology Press, 1984).
4. Fisher, R. A., Corbet, A. S. & Williams, C. B. *J. Anim. Ecol.* **12**, 42–58 (1943).
5. Efron, B. & Thisted, R. *Biometrika* **63**, 435–447 (1976).
6. Goodman, L. A. *Ann. math. Statist.* **20**, 572–579 (1949).
7. Derenzo, S. E. & Hildebrand, T. H. *Nucl. Instrum. Meth.* **69**, 287–292 (1969).

Evidence for mass loss from IRAS observations of classical Cepheids

Hugh Deasy

Dunsink Observatory, Dublin 15, Ireland

C. J. Butler

Armagh Observatory, Armagh, Northern Ireland BT61 9DG, UK

Mass loss from classical Cepheid variables is suspected on two grounds. First, it may provide an explanation of the persistent discrepancy between estimates of Cepheid masses based on the theories of stellar pulsation and of stellar evolution¹. Second, theoretical models of pulsating atmospheres² suggest that a pulsation mechanism may be responsible for causing, or enhancing, mass loss from Cepheids. In order to test these hypotheses, evidence of anomalous mass loss was sought by making a comparison between the infrared emission of Cepheids and that of nonvariable supergiants in the same luminosity and effective temperature range. A search of the IRAS (Infrared Astronomy Satellite) catalogue³ found a number of Cepheids and stable supergiants which showed

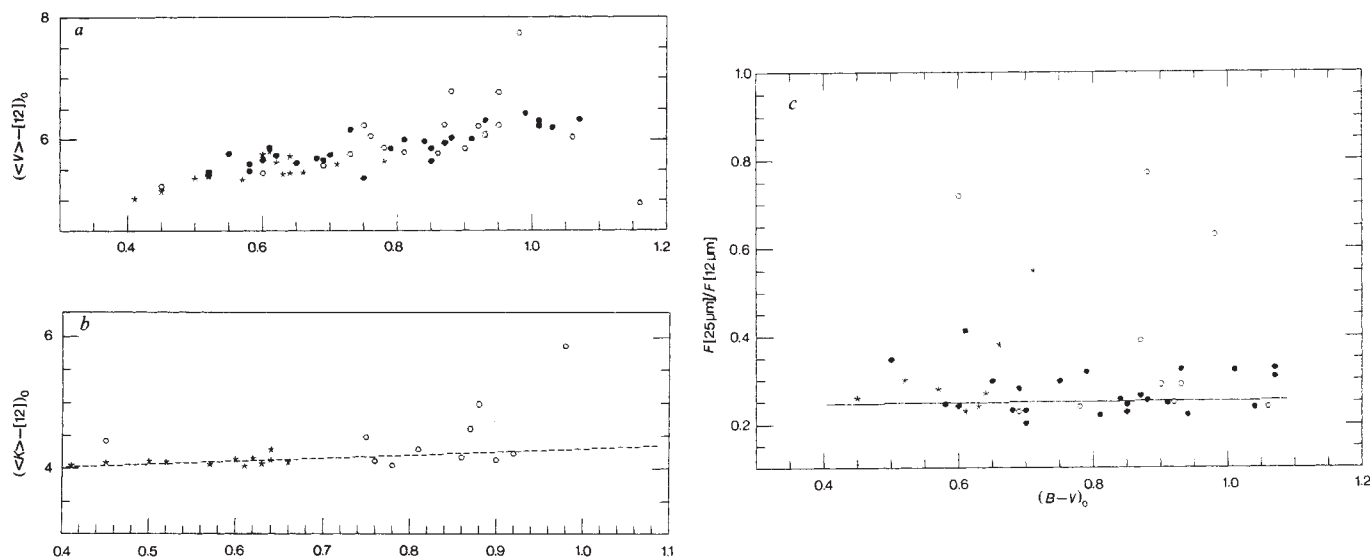


Fig. 1 a, Plot of intrinsic difference between V magnitude at mean intensity and IRAS 12-μm magnitude $\langle(V) - [12]\rangle_0$ versus intrinsic colour index $(B - V)_0$ for short-period Cepheids (asterisks), long-period Cepheids (open circles) and nonvariable supergiants (filled circles). b, Plot of intrinsic difference between K magnitude at mean intensity and IRAS 12-μm magnitude $\langle(K) - [12]\rangle_0$ versus $(B - V)_0$ for Cepheids. Symbols as in a; the line represents the locus of black-body values for the effective temperature range of the Cepheids. c, Ratio of flux at 25 μm to flux at 12 μm, $F(25\mu\text{m})/F(12\mu\text{m})$, plotted against $(B - V)_0$. Symbols as in a; the black-body locus is also shown.

emission in at least one of the IRAS wavelength bands. Some long-period Cepheids showed infrared excesses with respect to their non-pulsating counterparts, while emission from Cepheids with periods of less than 10 days was comparable to the levels seen in the stable supergiants. Mass loss rates of up to $7 \times 10^{-7} M_{\odot} \text{ yr}^{-1}$ were derived from the infrared excesses, which is sufficiently high to have a major effect on the evolution of these stars.

Cepheids and nonvariable supergiants were chosen for this study according to the following criteria: (1) The magnitude in the Johnson V band at mean intensity, $\langle V \rangle < 9.0$; (2) $-7.0 < \text{the absolute visual magnitude } M_V < -2.5$; (3) the colour excess $E(B - V) < 1.0$; and (4) $0.4 < \text{the intrinsic colour index } (B - V)_0 < 1.1$. A search of catalogues⁴⁻⁹ yielded a list of objects, 71 of which (33 Cepheids and 38 nonvariable supergiants) were found to be coincident with IRAS point sources. Of the Cepheids, all showed emission in the 12-μm band, 18 at 25 μm, 6 at 60 μm and one at 100 μm. To assist us in the comparison of the observed infrared emission with that expected from a black body, we determined the broad-band colour $\langle(V) - [12]\rangle_0$, where $[12]$ is the infrared magnitude defined as $[12] = -2.5 \log F(12\mu\text{m})$, where $F(12\mu\text{m})$ is the IRAS flux in Janskys in the 12-μm band. Similarly we determined $\langle(K) - [12]\rangle_0$, where $\langle K \rangle$ is the magnitude at mean intensity in the infrared band with mean wavelength 2.2 μm for Cepheids for which JHK photometry was available^{5,10}. As far as we could determine, no K measurements of the nonvariable supergiants have been published.

Figure 1a is a plot of de-reddened $\langle(V) - [12]\rangle_0$ against $(B - V)_0$ for the Cepheids and nonvariable supergiants. Whereas the short-period Cepheids are mingled with the supergiants, some of the longer-period and redder Cepheids have $\langle(V) - [12]\rangle_0$ values which are significantly larger than supergiant values for the same $(B - V)_0$. One 19-day Cepheid, RU Sct, has pronounced infrared excess with $\langle(V) - [12]\rangle_0 > 7$. Also, Fig. 1b shows that the short-period Cepheids fall along the locus of expected black-body values of de-reddened $\langle(K) - [12]\rangle_0$, whereas several of the long-period Cepheids lie above this curve. It should be noted that some of the long-period Cepheids with large values of $\langle(V) - [12]\rangle_0$ do not appear in Fig. 1b, as no K photometry is available. Figure 1c shows the ratio of fluxes in the 25-μm and 12-μm IRAS bands, $F(25\mu\text{m})/F(12\mu\text{m})$, for stars detected at 25 μm. Whereas all the supergiants and most

of the Cepheids fall around the expected black-body value of ~ 0.25 , some long-period Cepheids lie significantly above this level. All of the latter also lie above the supergiant locus in Fig. 1a, demonstrating the significance of the $\langle(V) - [12]\rangle_0$ excesses.

The fact that $\langle(V) - [12]\rangle_0$ for some long-period Cepheids exceeds that for the nonvariable supergiants points to a specifically pulsation-related phenomenon. Further support for this comes from considering the dependence of $\langle(V) - [12]\rangle_0$ on period, P (Fig. 2). Cepheids with the largest excesses with respect to the nonvariable supergiants have $\log_{10} P$ values between 1.2 and 1.6, with a peak at 1.3. The level of excess is also correlated with the amplitude of the light curve, again indicating a connection with the pulsation.

In addition to our general comparison with supergiants, we can look for evidence among Cepheids of an infrared excess relative to standard black-body curves. In Fig. 3 we show the observations for three Cepheids for which both JHK^{5,10} and IRAS³ photometry are available (based on broad-band photometry^{3,5,9-11} and spectrophotometry in the 0.3–0.6-μm region¹²), together with black-body radiation curves for their estimated effective temperatures.

For one Cepheid, δ Cep, the IRAS observations fit the black-body curve well. For the other two Cepheids, RS Pup and RU Sct, however, the IRAS fluxes show clear evidence of an infrared excess. RS Pup, the 41-day Cepheid, with its conspicuous sur-

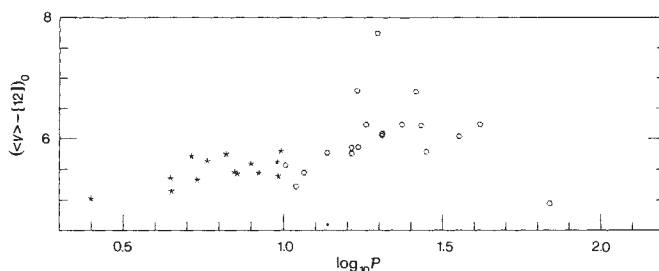


Fig. 2 Plot of intrinsic colour $\langle(V) - [12]\rangle_0$ (see Fig. 1a) against $\log_{10}$ of period for Cepheids. Symbols as in Fig. 1.

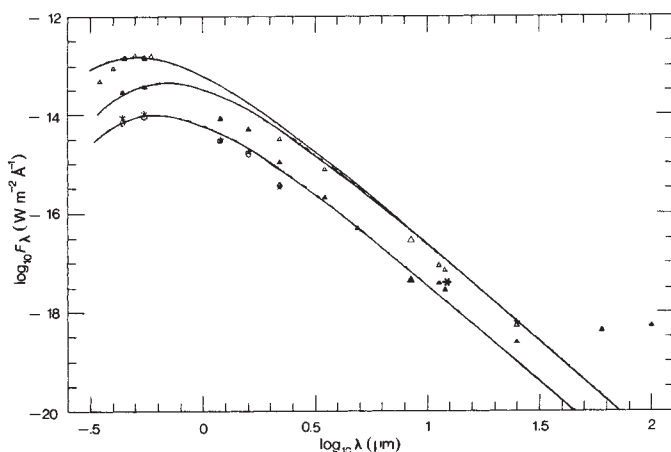


Fig. 3 Energy distributions (plots of flux against wavelength) for three Cepheids: δ Cep (Δ), RS Pup ($\blacktriangle$) and RU Sct (*). The broken lines represent black-body energy distributions for estimated effective temperatures of the stars.

rounding nebosity¹³, shows extended emission in the 25-, 60- and 100- μ m bands characteristic of a black-body emitter at around 40 K. The excess at 12 μ m in this star is, however, more likely to be due to non-planckian emission from inlying material.

There has recently been much discussion of the well-known discrepancy between estimates of Cepheid masses based on the theories of stellar evolution and pulsation. There is no consensus as to the reason for the discrepancy; efforts to date have centred on adjustments to the distance scale¹, and the use of inhomogeneous pulsation models featuring surface helium enrichment¹⁴ or tangled magnetic fields¹⁵. Recently it has been suggested² that pulsation-induced mass loss may be responsible for the consistently higher mass estimates based on evolutionary models (which neglect mass loss) relative to those from pulsation theory. To test this proposition, it is of interest to compare mass loss rates calculated from observational data with the rates needed to explain the mass discrepancy. Estimates of mass loss rate may be derived for Cepheids with significant infrared excesses from assuming that the excesses are due either to silicate emission from thin envelopes of circumstellar dust¹⁶ ($\dot{M}_{\text{dust}}$) or to free-free emission from hot ionized plasmas²³ ($\dot{M}_{\text{ff}}$). In the former case the radius of the inner dust shell is taken as the radial distance at which the radiative equilibrium temperature becomes low enough to allow dust condensation to take place.

Values of $\dot{M}_{\text{ff}}$ and $\dot{M}_{\text{dust}}$ for long-period Cepheids with significant excesses over the stable supergiants are given in Table 1. Upper limits on $\dot{M}_{\text{ff}}$ and $\dot{M}_{\text{dust}}$ are also given for some short-period Cepheids with relatively large values of $\langle V \rangle - [12]_0$. We see that the free-free estimates are higher than those derived assuming circumstellar dust. According to theoretical models of the atmospheres of pulsating stars (ref. 18 and L. A.

Willson, personal communication), the $\dot{M}_{\text{ff}}$ rates are more likely to apply to short-period Cepheids, or to any Cepheids with mass loss less than $10^{-6} M_{\odot} \text{ yr}^{-1}$, since shock heating is expected to be more efficient in heating the atmosphere in these cases, preventing the formation of dust grains. The derived $\dot{M}_{\text{ff}}$ rates, operating over typical Cepheid lifetimes of 10^3 – 10^6 yr, would produce cumulative mass losses of between $10^{-5} M_{\odot}$ and $1 M_{\odot}$, whereas if the emission is due to dust, the total mass lost would lie in the range $10^{-6} M_{\odot}$ to $0.1 M_{\odot}$. Thus in some cases the fraction of the initial Cepheid mass removed could be high enough to explain the mass discrepancy. In order to distinguish

Table 1 Estimates of mass loss rate for Cepheid variables with significant infrared excesses. Values in parentheses indicate data of lower significance, included for comparison purposes.

Name	Period (days)	$\dot{M}_{\text{ff}}$ ($10^{-8} M_{\odot} \text{ yr}^{-1}$)	$\dot{M}_{\text{dust}}$ ($10^{-8} M_{\odot} \text{ yr}^{-1}$)
V350 Sgr	5.154	(2.2)	(0.3)
RX Cam	7.912	2.8	1.1
S Mus	9.659	(1.5)	(0.2)
XX Cen	10.956	(3.8)	(1.1)
RW Cam	16.413	8.3	2.5
CD Cyg	17.071	20	3.7
RU Sct	19.698	70	8.0
X Pup	25.961	6.9	1.8
RY Vel	28.125	12	2.1
RS Pup	41.388	10	1.7

whether the infrared emission is free-free or due to circumstellar dust, we would require additional observations at shorter wavelengths.

A 'trapping' mechanism has been suggested² whereby mass loss due to pulsation may cause the tips of the blue loops of intermediate-mass stars in the HR diagram to migrate into the instability strip. The resultant increase in Cepheid lifetime would augment the cumulative mass loss. $\dot{M}_{\text{ff}}$ estimates derived for some Cepheids with periods of ~ 10 days are hardly high enough to cause trapping¹⁹. For longer-period Cepheids, mass loss rates somewhat higher than those observed would be required for this effect to occur.

Although recent work²⁰ implies that the mass discrepancy is most serious for Cepheids with periods of < 10 days, there is no evidence from infrared excesses for the implied high mass loss rates among stars in this period range. Evidence for the presence of stellar winds in short-period Cepheids comes from the net asymmetry with phase of ultraviolet resonance lines²¹, and from blueshifted absorption components. A mass loss rate of $\sim 10^{-10} M_{\odot} \text{ yr}^{-1}$ is indicated for the 10-day Cepheid ζ Gem¹⁷. Although insufficient to produce enough mass loss to explain the mass discrepancy, such low rates could produce the surface helium enrichment which has been suggested as a means of alleviating the most serious mass anomaly, namely that for short-period Cepheids simultaneously oscillating in two modes, or with bumps on their light curves²².

Received 22 November, 1985; accepted 17 February 1986.

- Cox, A. N. *A. Rev. Astr. Astrophys.* **18**, 15–41 (1980).
- Willson, L. A. & Bowen, G. *Nature* **312**, 429–431 (1985).
- IRAS Point Source Catalog (Joint IRAS Science Working Group, 1985).
- Fernie, J. D. & Hube, J. O. *Astr. J.* **73**, 492–499 (1968).
- Welch, D. L. *et al. Astrophys. J. Suppl.* **54**, 547–579 (1984).
- Schmidt, E. G. *Astrophys. J.* **172**, 679–683 (1972).
- Eggen, O. J. *Astr. J.* **90**, 1260–1277 (1985).
- Fernie, J. D. & Hube, J. O. *Astrophys. J.* **168**, 437–442 (1971).
- Schaltenbrand, R. & Tammann, G. A. *Astr. Astrophys. Suppl.* **4**, 265–314 (1971).
- McGonegal, R., McAlary, C. W., McLaren, R. A. & Madore, B. F. *Astrophys. J.* **269**, 641–644 (1983).

- Gehrz, R. D. *Astrophys. J.* **178**, 715–725 (1972).
- Patterson, R. S. & Neff, J. S. *Astrophys. J. Suppl.* **41**, 215–241 (1979).
- Havlen, R. J. *Astr. Astrophys.* **16**, 252–267 (1972).
- Cox, A. N. *Astrophys. J.* **229**, 212–222 (1979).
- Stothers, R. *Astrophys. J.* **234**, 257–261 (1979).
- Gehrz, R. D. & Woolf, N. J. *Astrophys. J.* **165**, 285–294 (1971).
- Deasy, H. & Wayman, P. A. *Ir. astr. J.* (in the press).
- Willson, L. A. & Bowen, G. in *3rd Trieste Workshop on the Relationship between Chromospheric/Coronal Heating and Mass Loss* (eds Stalio, R. & Zirker, J.) (SISSA-ISAS, Trieste, 1985).
- Brunish, W. M., Willson, L. A. & Becker, S. A. *Bull. Am. astr. Soc.* (in the press).
- Schmidt, E. G. *Astrophys. J.* **285**, 501–514 (1984).
- Schmidt, E. G. & Parsons, S. B. *Astrophys. J.* **279**, 202–212 (1984).
- Cox, A. N., Michaud, G. & Hodson, S. W. *Astrophys. J.* **222**, 621–626 (1978).
- Gehrz, R. D. & Woolf, N. J. *Astrophys. J.* **165**, 285–294 (1971).

More dark matter around Uranus and Neptune?

B. Sicardy*, F. Roques*, A. Brahic*†, P. Bouchet‡, J. P. Maillard§ & C. Perrier||

* Observatoire de Paris, 92190 Meudon, France

† Université Paris 7, 2 place Jussieu, 75005 Paris, France

‡ European Southern Observatory, Casilla 19001, Santiago de Chile 19, Chile

§ Canada France Hawaii Telescope Corporation, PO Box 1597, 95743 Hawaii, USA

|| Observatoire de Lyon, 69230 Saint Genis Laval, France

In addition to their natural satellites, the giant planets possess a surprisingly rich population of moonlets, rings, arcs and diffuse haloes of dust¹⁻⁶, the dynamics of which involves gravitational interactions⁷ and radiation and magnetic forces^{8,9}. Recent analyses (by A.B., W. Hubbard, F.R. and B.S.) of isolated events observed around Neptune indicate the existence of an incomplete or at least a highly azimuthally variable ring or 'arc' around Neptune. Similar phenomena may be detectable around Uranus. In order to look for yet undetected dark matter around Uranus and Neptune, the best signal-to-noise ratio stellar occultations have been reanalysed. To be confirmed, an 'event' should be observed with more than one telescope on the same site since the Fresnel scale (see below) is about 2–3 km at the level of Uranus and Neptune, and it should also be observed from widely separated observatories (100–1,000 km) to have an idea of the spatial extent of the occulting object, and to distinguish between satellites and ring-like structures. Moreover, theoretical models of diffracting bars or circular objects provide useful tests to reject some isolated events as being incompatible with real and celestial events¹⁰. We report here an investigation of the surroundings of Uranus and Neptune using stellar occultations observed at the European Southern Observatory (ESO) on 22 April 1982 and at the Canada France Hawaii Telescope (CFHT) on 15 June 1983, respectively. The high signal-to-noise ratio of these observations sets stringent limits on the presence of matter around these planets.

The stellar occultation technique makes it possible to detect material that would otherwise be invisible to classical ground-based direct imagery. It is complementary to space missions, such as Voyager 2 at Uranus in January 1986 and Voyager 2 at Neptune in August 1989. This technique is ideal for the study of narrow continuous rings at high spatial resolution. However, the occultation observations have their own limits and the presence of matter is probably not restricted to what is visible at first glance. Two main kinds of events are difficult to interpret without ambiguity in the occultation curves: First, sharp 'isolated events', that is dips observed on only one side of the planet and/or by only one telescope and/or during only one occultation; second, broad and faint absorptions, which are difficult to separate from transparency or seeing variations of the Earth atmosphere, or signal fluctuations due to slow tracking drift.

Table 1 lists the present observations^{11,12}. The 22 April 1982 occultation by Uranus and the 15 June 1983 occultation by Neptune were particularly interesting because the occulted stars were bright in the infrared. The Neptune occultation presented a unique opportunity for probing Neptune's equatorial plane

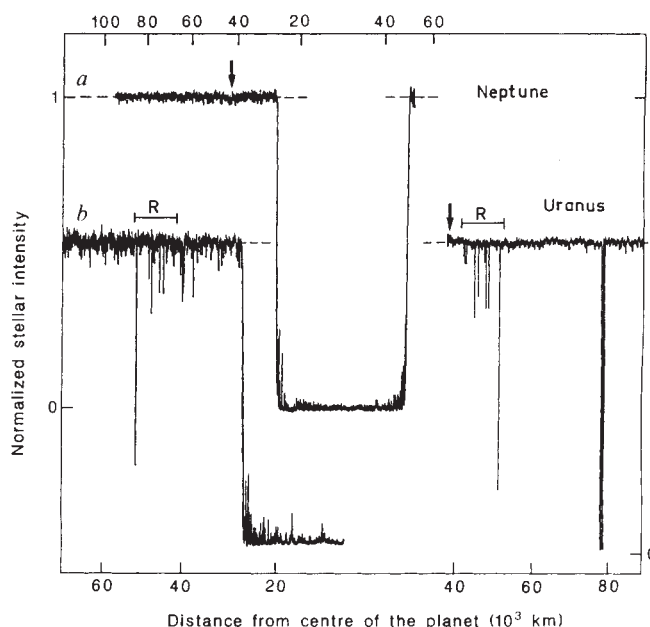


Fig. 1 *a*, Occultation of 15 June 1983 by Neptune, from the 3.6-m telescope of Canada France Hawaii observatory (CFHT, Hawaii) plotted at a time resolution of 0.3-s, from 13 h 45 min 56.7 s to 14 h 57 min 41.7 s UT. *b*, Occultation of 22 April 1982 by Uranus, from the 3.6-m telescope of the European Southern Observatory (ESO, La Silla), plotted at a time resolution of 0.315-s, from 01 h 14 min 35 s to 03 h 35 min 30 s UT. Both curves have been corrected for air-mass extinction and normalized to an out of atmosphere value (the levels unity, obtained by fitting all the data to a Bouguer curve are represented by the dashed lines). The bars labelled R locate Uranus rings and the arrows indicate small variations of signal discussed in the text. Drops of signal around 03 h 25 min 30 s UT in Uranus light curve are sky checks. The distance scales are given in the equatorial plane of respective planets.

down to the planet surface itself (latitude of the suboccultation immersion point, 0.8°).

The main limitation to the detection of faint material during stellar occultations is that, in typical conditions, there are signal fluctuations of the order of 1% because of sky transparency variations and seeing effects. During a few hours of observation, the signal is affected by the air-mass variation, and this must be taken into account when searching for faint variations of signal. So we have performed least-square fits using Bouguer's law to correct the data to an 'out of atmosphere' value. The remaining features in the curve are then due to atmospheric or instrumental effects (such as seeing, clouds, sensitivity baseline of the detector) and of course, to possible circumplanetary material.

Figure 1 shows the corrected light curve for the 22 April 1982 occultation by Uranus, observed by the 3.6-m ESO telescope. Due to larger air-mass, the pre-immersion noise is greater than that post-emersion. The interruption between 02 h 24 min and 02 h 48 min UT is the result of a temporary failure in the recording device. Except for a 1.5% enhancement of the signal just after 02 h 48 min UT (see arrow in Fig. 1), probably due to a recentring procedure, the average level of stellar intensity in the

Table 1 Conditions of observation

Date	Object	Site	Telescope	Detector	Band	Stellar magnitude	<i>B</i> * (deg)
22 April 1982	Uranus	ESO	3.6-m	InSb	K	5.2	74.61
			1.0-m	InSb	K		
15 June 1983	Neptune	CFHT	3.6-m	InSb	K	7.8	23.97

* *B* is the planetocentric declination of the Earth.

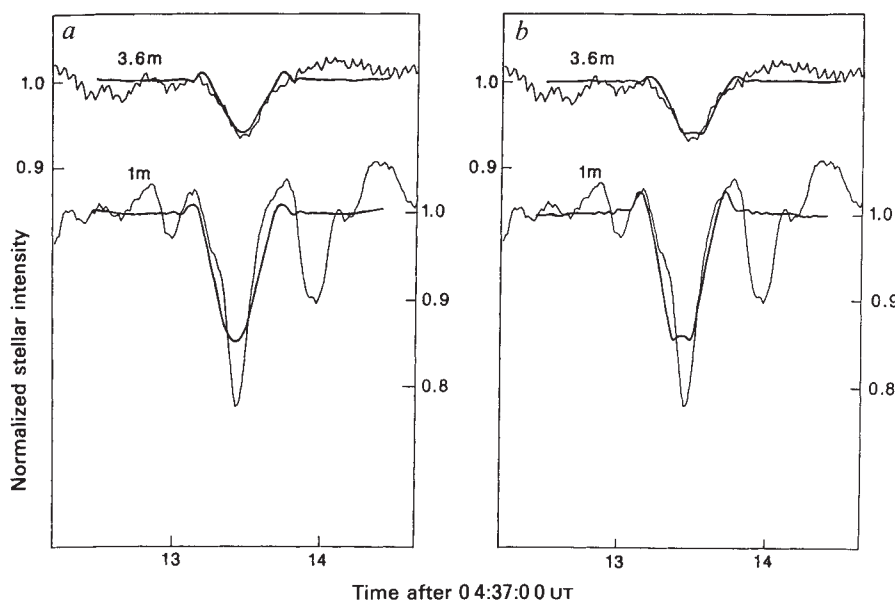


Fig. 2 The 22 April 1982 observation of a simultaneous event at the 3.6-m and 1-m telescopes of ESO, plotted in both cases at 0.015-s time resolution (thin lines). The heavy lines are results of model fitting: *a*, 3.6-m telescope: occultation by a 1.4-km diameter object with an impact parameter (with respect to the star) of 2.1 km 1-m telescope: fit with an object diameter of 2.5 km with and impact parameter of 1.8 km. *b*, 3.6-m telescope: occultation by a ring of apparent width 0.70 km and apparent transmission coefficient 0.55. 1-m telescope: fit with an apparent width 0.64 km and apparent transmission coefficient 0.04.

vicinity of the ring system does not exhibit any drop greater than 1%. This provides an upper limit of $\tau = 0.01$ for the normal optical depth of possible broad continuous equatorial rings of Uranus, around the nine narrow rings. Matthews *et al.*¹³ and Elliot and Nicholson³, based on infrared scannings of Uranus, independently give a lower limit of 0.003–0.01 for the optical depth of broad (5,000 km) uranian rings.

Figure 1 shows the corrected light curve from the occultation of 15 June 1983 by Neptune obtained at CFHT. The curve does not show any broad drop greater than 1.5% (see arrow in Fig. 1 around 14 h 13 min 10 s UT). Because of the tilt angle B between the line-of-sight and the Neptune's equatorial plane (see Table 1), this gives an upper limit of $\tau = 0.006$ for the normal optical depth of a possible broad continuous neptunian ring.

Diffraction and the finite size of the star set a fundamental limit to the detection of small objects. Diffraction smears out details smaller than the Fresnel scale $\sqrt{\lambda D}$, where λ is the wavelength of observation and D the distance to the planet. The finite stellar size smooths details smaller than the apparent stellar diameter at the planet distance. The Fresnel scale is ~ 2 –4 km in the near infrared at Uranus and Neptune distances and the stellar diameter ranges from 1 to 20 km, depending on the occulted star.

For objects smaller than the minimum spatial resolution, it is convenient to define an 'equivalent width' E (ref. 14), which is the width, in the equatorial plane of the planet, of a hypothetical opaque ring blocking as much light as the observed object.

$$E = W(1 - F) \quad (1)$$

where W is the actual width of the ring and F its normal transparency coefficient, which is assumed to be constant across the ring.

In addition to the nine ring occultations observed around Uranus on 22 April 1982, several isolated events were recorded on both the 1-m and the 3.6-m telescopes at ESO. Most of them are easily identified as being caused by electronics noise (they have a characteristic signature) or atmospheric turbulences (checked on the image of the guiding star). Among the remaining few dubious events, we eliminate those which are seen on only one light curve. One event appears simultaneously in the two recordings at 04 h 37 min 13.4 s UT (ref. 15). If this event corresponds to an object located in the equatorial plane of Uranus, its distance from the planet's centre is 157,900 km.

Although simultaneous, the events have different depths, the 1-m one being more than twice as deep as the 3.6-m event (Fig.

2). If the dips were caused by material around Uranus, two possibilities can *a priori* explain this difference: either the higher noise observed on the 1-m light curve (see, for instance, the dip half a second after the main event) caused a spurious enhancement of the event, or the occultation was caused by an isolated body which was 'cut' at two different chords by the apparent path of the star seen from each telescope. To test the last possibility, we have performed simulations of diffraction patterns due to spherical bodies of radius R_{obj} , passing at a minimum apparent distance P (impact parameter) of a star with apparent radius R_{stel} , the free parameter being R_{obj} and P (ref. 10).

Several reasonable fits can be obtained using different values of R_{obj} and P . Figure 2a shows an example of the fits which can be obtained with this model. The less noisy 3.6-m event is quite reasonably fitted by a 1.4-km diameter object occulting the star with an impact parameter of 2.1 km. On the other hand, the 1-m event is poorly fitted. Fitting the two events independently should give the same object radius, but different impact parameters, compatible with the separation of 0.17 km of the two telescopes, projected in the plane perpendicular to the line-of-sight. No solution with impact parameter difference < 1 km can be found. Moreover, the 1-m event is too narrow (2.7 km) and no solution can fit both the depth and width of this event. So, the 1-m event is either too noisy to be reliably fitted by any model and the 1-m curve provides information only about the time of the event, or the spherical moonlet hypothesis should be rejected.

Table 2 Results of model fittings for the 22 April 1982 event of ESO

Occultation by a isolated moonlet				
	Diameter (km)	Impact parameter (km)	Mid-time (UTC)	
3.6-m	1.4	2.1	04.37.13.445	
1.0-m	2.5	1.8	04.37.13.410	
Occultation by a semi-transparent ring				
	Width (km)	Transparency coefficient	Equivalent width (km)	Mid-time (UTC)
3.6-m	0.70	0.55	0.35	04.37.13.460
1.0-m	0.64	0.04	0.7	04.37.13.395

Those values assume an apparent stellar diameter at Uranus of 7.1 km (ref. 3).

Similar simulations of diffraction patterns produced by a fragment of a semi-transparent ring have been performed (Fig. 2b). The free parameters are the width W of the ring and the transmission f_0 . Several solutions are acceptable, corresponding to rings of equivalent width $E = 0.35 \pm 0.05$ km. Using equation (1), the range of the actual width lies between 0.7 and 4.0 km. The best fit for the 3.6-m event is obtained with $W = 0.7$ km and $f_0 = 0.55$. Again, the 1-m event is poorly fitted by a semi-transparent ring model. Its greater depth would require an equivalent width of at least 0.7 km. The timing of the events cannot bring any additional information because the expected lag between the dips should be ~ 0.04 s, which is of the same order of magnitude as the accuracy of the fits (~ 0.05 s).

The quasi-simultaneity of the two dips at the two telescopes and the good fits obtained with theoretical models for the less noisy supports the existence of the 3.6-m event. In contrast, there are some doubts because the 1-m event is too narrow and too deep with respect to the 3.6-m event. However, the lower quality of the 1-m signal means that this difference could be due to noise only.

If this event corresponds to a 'real' uranian object, then it could be either a 2–4-km sized moonlet having partially occulted the star, or a narrow ring-like structure of equivalent width 0.35 km at 157,900 km from the centre of Uranus, that is, between the orbit of Miranda (130,000 km) and Ariel (192,000 km). Simultaneous observations made at the Cerro Tololo Inter-American Observatory (CTIO), located 100 km south from ESO give a strong constraint on the azimuthal extension of such a structure. The CTIO light curve does not show any simultaneous event at the level presented here (J. L. Elliot, personal communication). Such an object, which is particularly difficult to detect using ground-based techniques, might be observed on the Voyager 2 spacecraft images in forwards scattered light if it is associated with dust.

The 15 June 1983 Neptune occultation observed from the CFH telescope did not show any isolated event. The peak-to-noise indicates that no sharp event deeper than 2.5% was detected. With the rather large (~ 1.8 s) overall time constant of the data acquisition chain, the upper limit for the equivalent depth of any occulting narrow rings is 0.35 km on the pre-immersion part of the occultation curve, corresponding to a distance between 26,000 and 92,000 km in the Neptune equatorial plane (between 1 and 3.5 neptunian radii). The short recording made after the emersion (Fig. 1) corresponds to a probe of the Neptune equatorial plane between 50,100 and 52,200 km. No obvious sharp event has been detected on this part of the light curve.

The particularly high signal-to-noise ratio of the observations presented above allows us to draw several conclusions.

(1) An upper limit of respectively 0.01 and 0.006 can be set for the normal optical depth of any continuous diffuse equatorial rings of Uranus and Neptune. A similar analysis by Elliot *et al.*¹⁴ of other observations of the 15 June 1983 Neptune occultation gives an upper limit of 0.016 for Neptune surroundings. Compared with the faint jovian rings of optical depth $< 5 \times 10^{-5}$ (ref. 4), these limits are still high. Faint uranian and neptunian rings can thus escape detection by such ground-based occultation observations. Furthermore, the transient presence of material around these planets cannot be excluded and one can ask if the rings may exhibit short timescale dynamical processes.

(2) From the first analysis, the presence of nine continuous well-defined rings around Uranus may have concealed the presence of arcs similar to those observed around Neptune^{6,21}. The event we have reported is typical of such dubious dips which require careful analysis.

(3) Uranus may possess a small kilometre-sized moonlet or a 0.35 km large ring-like structure at a distance of 157,900 km from its centre.

(4) No matter has been detected around Neptune on 15 June 1983²⁰. This is interesting when compared with 'arc' detections around the planet^{6,16,21} at $66,000 \pm 3,000$ km (22 July 1984 detec-

tion⁶) or around 54,000 km (20 August 1985 detection at Mauna Kea site from the Infrared Telescope Facility, IRTF, and the CFHT^{17,18}) of around 75,000 km (24 May 1981¹⁹).

This 15 June 1983 observation sets severe constraints on the continuity of neptunian arcs. Even if inclined or eccentric, these arcs are highly variable in azimuth. If they are continuous, they must be either narrower than about 300 m or with a normal optical depth < 0.006 . If not, they should have been detected on 15 June 1983.

Constructive and willingful cooperation programs for future observations are urgently needed for a better understanding of surroundings of Uranus and Neptune²².

Part of this work was supported by the European Southern Observatory, the Canada France Hawaii Telescope Corporation, the University of Paris, Observatory of Paris, INAG, INSU and the Centre de la Recherche Scientifique. C.P. was on a Fellowship contract with ESO at Garching at the moment of observation. We thank D. Cruikshank and E. Becklin for assistance during the 15 June 1983 observation at CFHT.

Received 22 January; accepted 13 February 1986.

- Owen, T. *et al.* *Nature* **281**, 442–446 (1979).
- Smith, B. A. *et al.* *Science* **212**, 163–191 (1981).
- Elliot, J. L. & Nicholson, P. D. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 25–72 (University of Arizona Press, Tucson, 1984).
- Burns, J. A., Showalter, M. R. & Morfill, G. E. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 200–272 (University of Arizona Press, Tucson, 1984).
- Showalter, M., Burns, J. & Cuzzi, J. *Nature* **316**, 526 (1985).
- Brahic, A. *et al.* *Nature* (Submitted).
- Goldreich, P. & Tremaine, S. A. *Rev. Astr. Astrophys.* **20**, 249–283 (1982).
- Mignard, F., in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 333–366 (University of Arizona, Tucson, 1984).
- Burns, J. A., in *Formation des Systèmes Planétaires* (ed Brahic, A.) 472–442 (Centre National d'Etudes Spatiales, Cépadués Editions, Toulouse, 1982).
- Rouques, F., Moncuquet, M., Sicardy, B. & Brahic, A. *Astron. J.* (in the press).
- Klemola, A. R., Mink, D. J. & Elliot, J. L. *Astr. J.* **86**, 138–140 (1981).
- Mink, D. J., Klemola, A. R. & Elliot, J. L. *Astr. J.* **86**, 135–137 (1981).
- Matthews, K., Neugebauer, G. & Nicholson, P. D. *Icarus* **52**, 126–135 (1982).
- Elliot, J. L., French, R. G., Meech, K. J. & Elias, J. H. *Astr. J.* **89**, 1587–1603 (1984).
- Sicardy, B. *et al.* in *Les Anneaux des Planètes Proc. IAU Colloq. No. 75*, 181–191 (1984).
- Hubbard, W. B. *Science* **231**, 1276–1278 (1986).
- Cooke, M. L., Nicholson, P. D., Matthews, K. & Elias, J. H. *Bull. Am. astr. Soc.* **17**, 719 (1985).
- Sicardy, B. *et al.* *Bull. Am. astr. Soc.* **17**, 923 (1985).
- Reitsema, H. J., Hubbard, W. B., Lebofsky, L. A. & Tholen, D. J. *Science* **215**, 289–291 (1981).
- Sicardy, B. *et al.* *Bull. Am. astr. Soc.* **15**, 816 (1983).
- Hubbard, W. B. *et al.* *Nature* **319**, 636–640 (1986).
- Maddox, J. *Nature* **318**, 505 (1985).

Fractal viscous fingering in clay slurries

Henri Van Damme, François Obrecht, Pierre Levitz, L. Gataineau & Claude Laroche

Centre de Recherche sur les Solides à Organisation Cristalline Imparfaites, CNRS, 45071 Orleans Cedex 2, France

The fractal growth of viscous fingers on injection of a low-viscosity fluid into a high-viscosity fluid represents a type of hydrodynamic instability which was first demonstrated by Nittman *et al.*^{1,2}, by pushing water through an aqueous polymer solution. The conditions for the occurrence of the fractal morphology (in place of classical smooth fingering of the Saffman–Taylor³ type) are a combination of high viscosity contrast and low interfacial tension, which leads to a high capillary number and, most probably, non-newtonian rheological properties^{2,4}. In fact, these are conditions which are easily fulfilled on injecting a solvent into a concentrated suspension of colloidal particles in the same solvent. In particular, we show here that fractal viscous fingering generally occurs when water flows through clay slurries in a radial Hele–Shaw cell, but the fractal dimension of the pattern is strongly dependent on the concentration of the clay slurry. Given the importance of clay/water systems in nature, this result has far-reaching consequences, which we discuss.

The clay used in this study was a ground crude bentonite from Wyoming which from a mineralogical standpoint is essentially a Na-montmorillonite. Bentonite–water systems display

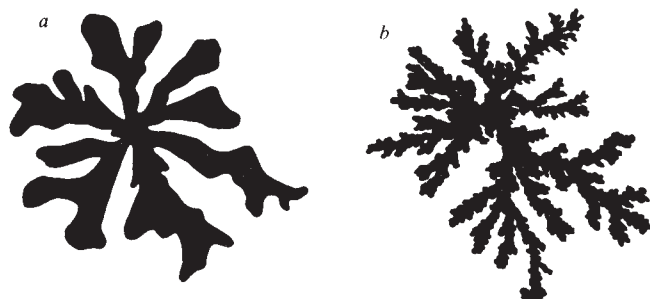


Fig. 1 Viscous finger patterns created by water in aqueous clay slurries containing 10% (a) and 7% (b) by weight of bentonite. The pattern in both a and b is ~ 0.4 m wide at its widest point.

an extremely broad range of mechanical properties, from very fluid suspensions to plastic pastes and finally to brittle solids. Here we are interested in the concentrated suspensions and slurries ($0.05 < \text{clay/water} < 0.15$, w/w). In this concentration range, bentonite slurries are highly viscous non-newtonian fluids with a strong thixotropic and plastic character⁵. Their interfacial tension with pure water is expected to be low because the interface between an aqueous clay slurry and water is essentially a water-water interface. Indeed, NMR and calorimetric measurements⁶ have shown that the relaxation rate of water in clay suspensions is not perturbed beyond a distance of $10 \text{ \AA}$ from the surface of the solid particles.

The experiments were performed in radial geometry, in a square Hele-Shaw cell (two closely spaced glass plates) of dimensions $0.5 \times 0.5 \times 0.0003$ m. Pure distilled water was injected (using a syringe needle) through a small hole at the centre of the bottom plate. The injection rate was between 1 and 2 ml s^{-1} . A variety of clay slurries with clay/water ratios in the range 4–15% (w/w) were prepared. However, experiments at clay/water ratios $> 10\%$ were hardly ever possible because of the difficulty of filling the cell. At the other extreme, well-

contrasted fingers (that is, sharp water-slurry interfaces) were difficult to obtain in the lower concentration range because of too-rapid mixing of the slurry with the water. Hence, experiments were restricted to slurries with clay/water ratios in the range 0.06–0.10 (w/w). The rheological properties of these shear-thinning slurries are summarized in Table 1.

Table 1 Rheological properties of the clay slurries

Clay/water ratio (w/w)	0.06	0.07	0.10
η_0 (cP)	320	1,160	5,700
η_∞ (cP)	80	325	750
σ_0 (dyne cm^{-2})	40	95	1,500

η_0 , The viscosity calculated by extrapolation of the shear stress/shear rate diagram to zero shear rate. η_∞ , The plastic viscosity, that is, the viscosity at infinite shear rate, in the linear part of the shear stress/shear rate diagram. σ_0 , Yield stress.

Figure 1a shows that in concentrated slurries (clay/water = 0.10) the pattern observed is compact. The fingers are wide and only weakly branched. As the concentration of the slurry diminishes, however, the fingers become narrower and the pattern displays a more ramified morphology (clay/water = 0.07; Fig. 1b). This tendency is reinforced at a clay/water ratio of 0.06 (Fig. 2); at this point, all the fingers are of approximately the same width.

The fractal dimensions of the viscous finger patterns were calculated from the photographs. The area covered by pure water within a circle of radius r , $S(r)$, was measured with a planimeter and plotted against r in log-log plots (see, for example, Fig. 3). Good power laws were obtained in all cases over at least one order of magnitude in r , showing that the distribution of water in the fingers follows a simple law of the type

$$M_w \propto r^D \quad (1)$$

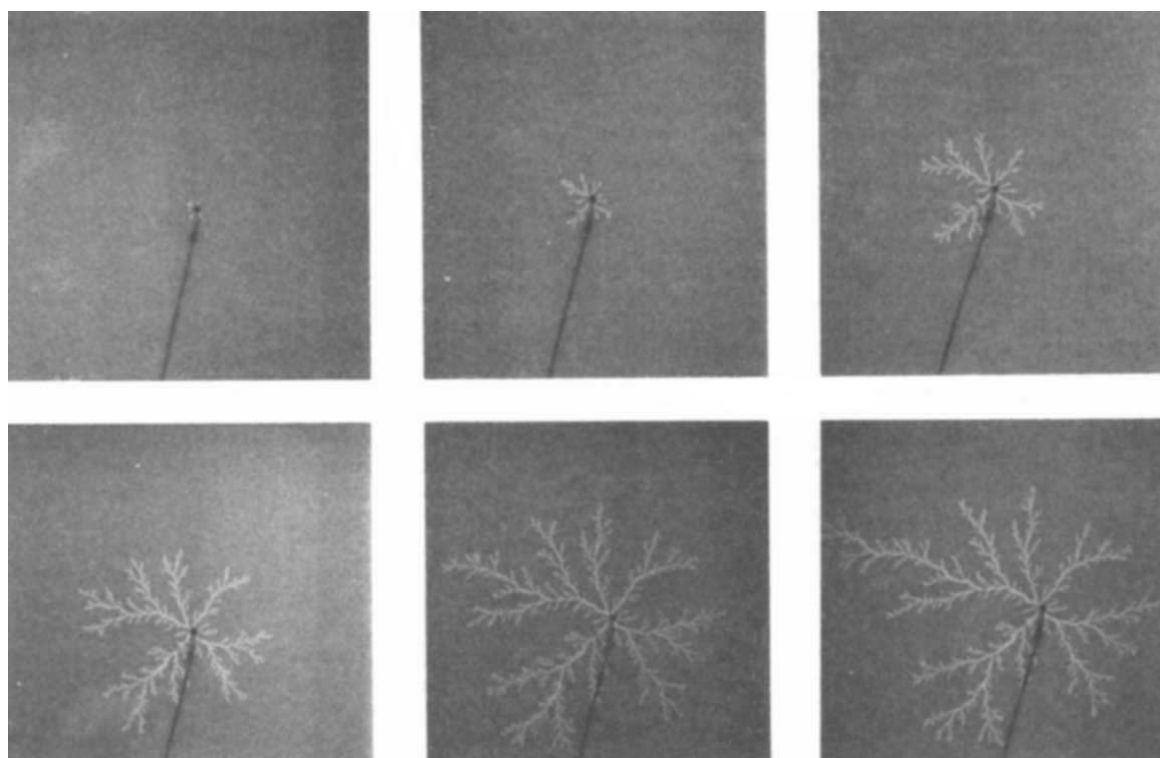


Fig. 2 Evolution of the viscous finger patterns created by water in an aqueous clay slurry containing 6% by weight of bentonite. The photographs are all at the same scale; the final pattern is ~ 0.4 ml wide.

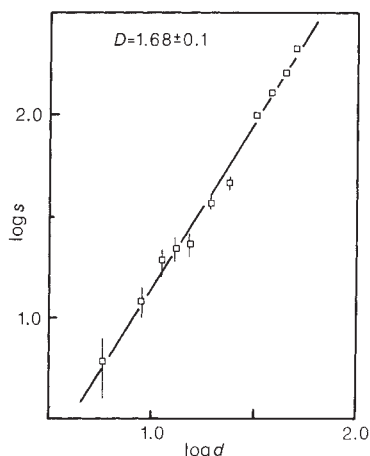


Fig. 3 Determination of the fractal dimension, D , of the viscous finger pattern in Fig. 2, from a log-log plot of $S(r)$, the surface area covered by pure water within a circle of radius r , against r .

where M_w is the mass of pure water. In other words, incipient fractal behaviour is present in all cases. The fractal dimension, D , was 1.84, 1.76 and 1.68 for clay/water = 0.10, 0.07 and 0.06 (w/w), respectively.

As pointed out by several authors^{1-4,7-10}, one of the most interesting aspects of fractal viscous fingering is its close relationship with other chaotic growth processes such as electrical breakdown of insulating media and diffusion-limited aggregation (DLA) of particles¹¹. A DLA cluster is constructed by releasing individual random walkers in a box containing a seed particle. In the simplest version ('regular' DLA) only one walker is released at a time, and the walkers stick as soon as they meet the central aggregate (or the seed particle, for the first walker). Regular DLA aggregates are self-similar objects with a fractal dimension $D = 1.68 \pm 0.05$ (ref. 12) when the embedding space (the box) is two-dimensional.

The close relationship between fractal viscous fingering dielectric breakdown and DLA stems from their connection with potential theory⁹. In all three cases, the behaviour of the system is governed by a Laplace equation, $\nabla^2 \phi = 0$, where ϕ is either the hydrodynamic potential in the high-viscosity fluid ($\phi = -p$, where p is the pressure), the electrostatic potential, or the particle concentration. In this framework, DLA appears as the prototype of purely stochastic growth. Kadanoff¹⁰ has shown that regular DLA corresponds to viscous fingering at the zero-surface-tension and infinite-viscosity-ratio limit. In bentonites, we obtained $D = 1.68$ at a clay/water ratio of 0.06 (Figs 2, 3).

Non-zero surface tension, which in Kadanoff's generalized DLA model can be accounted for by allowing for some inter-finger diffusion, leads to thicker fingers and a smoother boundary. A finite viscosity ratio leads to similar modifications, as shown by Daccord *et al.*² using a gradient-governed growth (GGG) model, in which the growth of the interface along the local pressure gradient is calculated by solving the general equation for the potential in the Hele-Shaw cell:

$$\nabla \cdot [(1/\eta) \nabla \phi] = 0 \quad (2)$$

both in the low- and the high-viscosity fluid. The smoother shape and the higher fractal exponent of the viscous fingers obtained with the more concentrated slurries can hardly be assigned to a finite viscosity effect. Rather, we suspect some interfacial energy effect, associated with the re-orientation of the colloidal particles along the moving interface.

The experiments reported here suggest that many of the branched structures observed in clay-water systems (river deltas, for example) are fractals generated by hydrodynamic instabilities at the water/clay slurry interface. Such structures are also relevant to the problem of nuclear waste disposal, where swelling clays have been considered as barrier materials¹³; frac-

tal viscous fingering could be a danger in this respect as it could allow the transport of solutions over large distances without giving the clay the opportunity to adsorb the ions. Our results suggest that little danger is to be expected before the clay has taken up an enormous amount of water. However, considering the extreme sensitivity of the surface and rheological properties of clays to many physicochemical factors such as the ionic strength of the medium⁵, this conclusion should not be applied too liberally.

Received 25 November 1985; accepted 26 February 1986.

1. Nittman, J., Daccord, G. & Stanley, H. E. *Nature* **314**, 141-144 (1985).
2. Daccord, G., Nittman, J. & Stanley, H. E. *Springer Proc. Phys.* **5**, 245 (1985).
3. Saffman, P. G. & Taylor, G. I. *Proc. R. Soc. A* **245**, 312 (1958).
4. Robinson, A. L., *Science* **228**, 1077-1080 (1985).
5. Van Olphen, H. *An Introduction to Clay Colloid Chemistry* (Wiley, New York, 1977).
6. Fripiat, J. J., Cases, J., Francois, M. & Letellier, M. *J. Colloid Interface Sci.* **89**, 378 (1982).
7. Sander, L. M. *Nature* **314**, 405-406 (1985).
8. Paterson, L. *Phys. Rev. Lett.* **52**, 1621-1624 (1984).
9. Witten, T. A. *Springer Proc. Phys.* **5**, 212 (1985).
10. Kadanoff, L. *J. statist. Phys.* **39**, 267 (1985).
11. Witten, T. A., & Sander, L. M. *Phys. Rev. Lett.* **47**, 1400-1414 (1981); *Phys. Rev.* **827**, 5685 (1983).
12. Meakin, P. *Phys. Rev. A* **27**, 1495 (1983).
13. Dayal, R. & Wilke, R. J. *Developments in Sedimentology* Vol. 35 (eds Van Olphen, H. & Veniale, F.) 771-787 (Elsevier, Amsterdam, 1982).

Volcanic shards from Santorini (Upper Minoan ash) in the Nile Delta, Egypt

Daniel Jean Stanley & Harrison Sheng

Division of Sedimentology, Smithsonian Institution, Washington, DC 20560, USA

One of the largest volcanic outbursts recorded in early human history is the cataclysmic eruption of Santorini (Thera) which occurred about 35 centuries ago in the southern Aegean Sea. This explosive event, termed Upper Minoan (or Upper Thera), produced a tephra-fall believed by archaeologists and volcanologists to have been widespread in the eastern Mediterranean¹⁻⁴. Upper Minoan ash has been found on land in the Aegean and Hellenic Arc regions including Crete^{5,6} and Rhodes⁷. To date, however, there has been no confirmed observation of this ash on land across the Mediterranean in North Africa. We now report evidence for the presence of ash ejected from this explosion in sediment cores recovered in the eastern Nile Delta of Egypt. This discovery of Upper Minoan glass shards extends considerably the zone of confirmed ash fall to the southeast of Santorini, and serves as evidence of a major natural phenomenon affecting Egypt during the reign of Pharaohs in the Eighteenth Dynasty. It is quite possible that this ash-fall event may have given rise to texts pertaining to darkening of the sky and veiling of the Sun in early historic records.

Volcanological investigations at Santorini have shown that the eruption which produced an enormous amount of pumice and ash about 3,500 years before present (yr BP) involved several phases⁸, and that the sequence included a Plinian pumice fall deposit, interbedded Surtseyan-type ash fall and base surge deposits, mudflow strata and ignimbrite interbedded with flood deposits⁹. The volume of tephra from this event is estimated at between 13 and 18 km³ (ref. 10). Waterborne pumice fragments of possible Santorini derivation have been found in widespread coastal regions, including the Peloponnese¹¹, Cyprus¹², Syria¹³ and Israel^{14,15}. Much of this pumice was displaced by flotation for considerable distances by the large counter-clockwise surface current circulation patterns in the eastern Mediterranean (compare ref. 16). The finer, less obvious, airborne volcanic glass shards settling from the Upper Minoan ash clouds form discrete layers identified in numerous deep-sea cores, but these occur in a considerably more restricted zone extending east and southeast

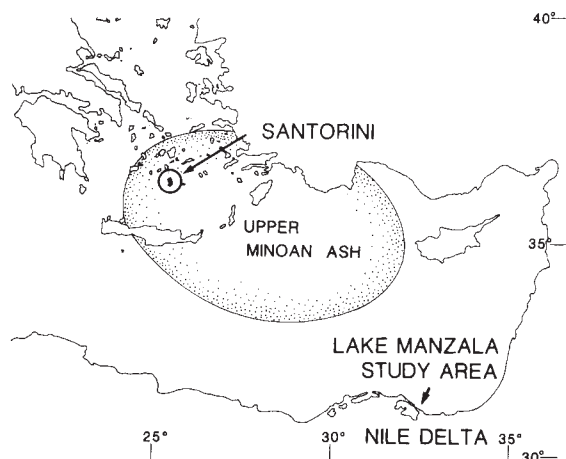


Fig. 1 Map of the eastern Mediterranean showing the eastern Nile Delta study area relative to Santorini. The predominant east-southeast distribution pattern of Upper Minoan ash²¹ is determined largely from deep sea cores (after refs 10, 17, 18, 21).

of Santorini and Crete and south of Turkey (Fig. 1). Identification of this tephra at sea is usually based on the index of refraction and chemical composition of volcanic glass and the age of the volcanic layers¹⁷⁻²².

It is thought that ash clouds from the Upper Minoan event could have reached eastern North Africa and the adjacent Near East on the basis of (1) the powerful nature of the eruption, (2) large volumes of ejected ash, (3) southeast trend of the ash clouds as determined from tephra layers at sea, and (4) relatively small distance between Santorini and the Egyptian coast (only about 800 km to the Nile Delta). Fall-out of this ash over Egypt has been predicted both by geologists^{17,23,24} and by archaeologists, usually as part of discussions emphasizing the probable destructive effects of this major explosive event over large parts of the eastern Mediterranean^{1,25}. More specific consideration of empirically derived relationships between eruption intensity, cloud height and distance^{26,27} indicates that volcanic ash particles as large as coarse silt (to over 40 μm) could have been carried across the Mediterranean by Upper Minoan ash clouds. Upper Minoan ash, however, has not been recovered in cores on the wide continental margin off Egypt and the Sinai²⁸ nor off Israel. The absence of ash layers here is probably the result of a masking effect by the transport of large volumes of terrigenous sediments from the Nile and their rapid deposition on these margins, and also by reworking of ash layers by benthic organisms on the seafloor. Sparks and others²⁹, for example, show that a discrete ash layer must be at least 0.62 cm thick to survive bioturbation and not be dispersed throughout the surrounding sediment. That volcanic shards from Santorini have not been recovered in North African and Levant deserts and archaeological sites is likely due to the subsequent reworking of thin ash layers by wind and dissemination of the glass particles.

As part of a large-scale Smithsonian-Egyptian project initiated in 1983 to interpret the Holocene evolution of the northeastern Nile Delta five cores recovered around Lake Manzala east of the Damietta Branch of the Nile (Fig. 2) were examined for possible ash. These cores, from 27 to over 35 m in length, consist largely of mud, with interstratified sand layers primarily at the top and base of the cores (described in ref. 30). Selected core samples were dated to establish a chrono-stratigraphic base for the Holocene section in this region, and mineralogical analyses were made of the silt and clay as well as of the sand fractions. Foraminiferal assemblages³¹ indicate that the mud sections 5 m and deeper below the core tops accumulated in shallow marine (<15 m deep) delta-front environments. Samples collected downcore at about a 1 metre interval revealed no volcanic

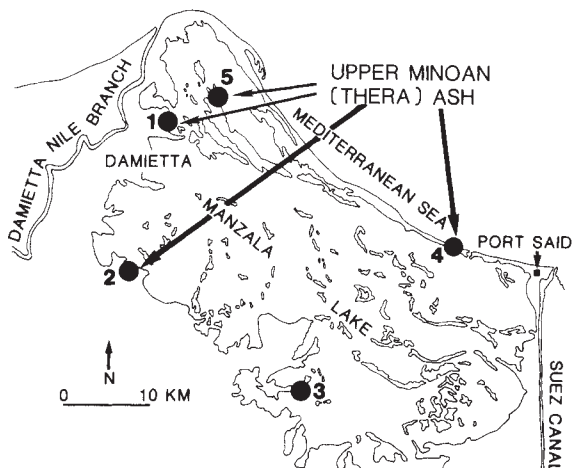


Fig. 2 Map showing the location of five sediment borings examined in the eastern Nile Delta.

fragments in the sand or coarser fractions. However, careful examination of the coarse silt fraction (38 to 63 μm) with a petrographic microscope during many months revealed the presence of glass shards concentrated between a depth of 5 to 7 m in four of the five cores.

About 40 volcanic shards (most ranging between 20 and 60 μm) were collected at this level. It is of note that this size coincides with the median diameter (20–40 μm) of Minoan ash fall deposits recovered in deep-sea cores, where the fine population is the same approximate size regardless of the distance from Santorini²⁹. These grains are somewhat more abundant in cores 2 and 4 than in cores 1 and 5 yet represent an infinitesimal part of the deposit (usually only about 1 shard per 10,000 to 20,000 grains) as a probable result of bioturbation and dispersal of the former ash layer. Two dates provide some basis for dating the levels in which the shards were found in core 2: one at 2,810 $\pm$ 70 yr BP (SI-No. 6467) at 5 m and another at 4,380 $\pm$ 95 yr BP (SI-No. 6475) at 7 m. From these two, an interpolated date of 3,595 years BP and interpolated minimal date of 3,512 years BP are calculated for sections at about 6 m below the core tops. There are problems in interpreting the actual timing of the Upper Minoan explosive events from radiocarbon dates collected at Santorini³²⁻³⁴, but most volcanologists view the eruption as having occurred sometime between 3,600 and 3,400 yr BP⁸.

The index of refraction of 30 grains ranges from 1.507 to 1.510, values comparable to published records of Upper Minoan

Table 1 Composition of volcanic glass shards from Thera Island (Minoan ash fall pumice) and Lake Manzala cores 2 and 4 based on microprobe analysis

	Minoan tephra layer Thera Island†	Core 2‡	Lake Manzala Core 4		
SiO ₂	73.2	68.4	70.2	71.2	71.2
Al ₂ O ₃	12.7	13.2	11.1	13.8	13.3
FeO*	1.7	1.5	3.1	1.9	3.3
MgO	0.3	0.1	0.3	0.3	0.4
CaO	1.2	0.5	0.4	1.3	1.6
K ₂ O	3.2	2.3	0.8	0.8	1.7
Na ₂ O	0.4	0.6	0.1	2.1	0.4
TiO ₂	0.28	0.2	0.16	0.27	0.25
MnO	0.02	0.04	0.04	0.04	0.10
Total	93.0	86.8	86.2	91.7	92.3

Total values are less than 100% due to several factors: the vesicular nature and possible water content of the grains, and the non-polished surfaces of some small silt-sized particles analysed by microprobe.

* Total iron as Fe.

† Average of data from two samples.

‡ Average of data from three samples.

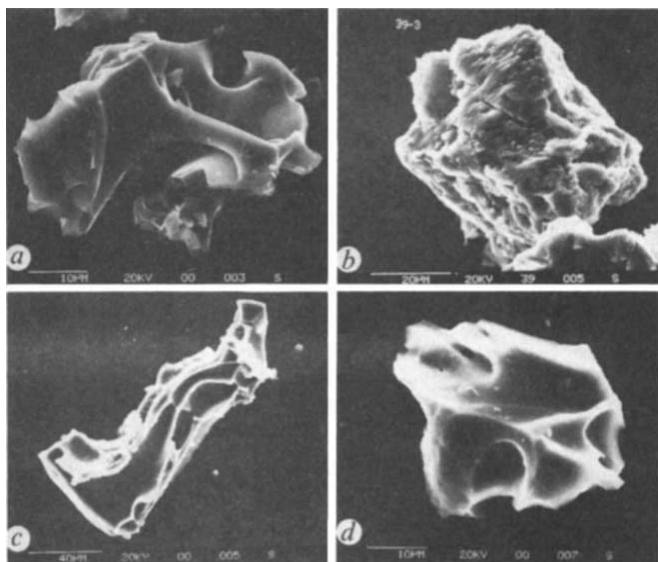


Fig. 3 SEM photomicrographs showing volcanic glass shards from a, Santorini (upper ash flow layer); b, Lake Manzala core 2, about 6 m depth; c and d, Lake Manzala core 4, about 6 m depth.

ash indices from Santorini¹⁹. After having established that these glass shards are restricted to a fairly specific depth-range in the cores, that the core sequences time-wise correspond reasonably with the Upper Minoan event, and that the grains are of a size and have an index of refraction comparable with ash recovered from Santorini, on Crete^{5,6} and at sea^{17,18,21}, we then selected 30 grains for scanning electron microscope (SEM) analysis. SEM photomicrographs showed that about 20 of these are angular and highly vesicular (Fig. 3b–d), typical of fine-grained tephra and similar to Upper Minoan ash collected at Akrotiri on Santorini (Fig. 3a). Concurrent examination of these same grains by energy dispersive X-ray analysis, which provides qualitative elemental data, confirms that these shards are volcanic glass. Electron microprobe (SEM-Q-ARL) analysis of six grains (6 to 12 analyses per grain) at accelerating voltage of 15 kilovolts and beam current of 0.15 microamps, provided data on nine major element oxides. Results were corrected according to Bence-Aldee procedures. Compositional data from six Lake Manzala core 2 and 4 samples are compared to the two Santorini Island Upper Minoan samples we also analyzed by microprobe (Table 1). These chemical data are similar to those of ash from Santorini and deep-sea cores published by others^{18–21}. It is of note, for example, that the TiO_2 values we measured are comparable to those of Upper Minoan ash collected at Santorini and in deep-sea cores²².

Although no single observation can by itself confirm that the volcanic shards we collected in the Nile Delta are of Upper Minoan origin, the correspondence of age, size, index of refraction and chemical composition, together, very strongly support a Santorini derivation. This first record of the ash in Egypt would substantiate the prediction of Van Bemmelen (ref. 24), p. 44) that Minoan ash might eventually be found on the southeastern shores of the Mediterranean 'in deposits of the lakes along the Suez Canal and in lakes along the lower Nile'. The ramifications of this discovery may prove significant with regards to interpreting early historical records pertaining to Egypt during the reign of Pharaohs of the Eighteenth Dynasty. The presence of Santorini volcanic ash provides a powerful non-archaeological argument that lends further support favouring an important natural phenomenon as recorded in diverse early documents. Of particular note are the 'plague of darkness' in the Biblical book of Exodus ('there may be darkness over the land of Egypt, even darkness which may be felt', Ex. 10:21) and citations with respect to veiling of the sun in several Egyptian papyrus records^{17,23–25,35,36}.

We thank Dr O. E. Frihy, Coastal Research Institute, Alexandria, for providing us with the suite of Lake Manzala core samples. At the Smithsonian Institution we thank Mr J. Nelen and Ms. D. Ross, Department of Mineral Sciences, for assistance with microprobe analyses, Mr W. Brown and Ms S. Braden, National Museum of Natural History, for help with the SEM study, and Dr R. Stuckenrath, Environmental Research Center, for radiocarbon dates. The manuscript was reviewed by Drs D. and C. Vitaliano. The study, part of the Mediterranean Basin (MEDIBA) Project, was funded by Smithsonian Scholarly Studies grants 1233S-405 and -502.

Received 20 December 1985; accepted 25 February 1986.

1. Marinatos, S. *Antiquity* **XIII**, 425–439 (1939).
2. Marinatos, S. (ed.) *Acta First int. Congr. on the Volcano of Thera, 1969* (Archaeological Services of Greece, Athens, 1971).
3. Doumas, C. (ed.) *Thera and the Aegean World I* (Aegean World, London, 1978).
4. Doumas, C. (ed.) *Thera and the Aegean World II* (Aegean World, London, 1980).
5. Vitaliano, C. J. & Vitaliano, D. B. *Am. J. Archeol.* **78**, 19–25 (1974).
6. Cadogan, G. & Harrison, R. K. in *Thera and the Aegean World I* (ed. Doumas, C.) 235–255 (Aegean World, London, 1978).
7. Doumas, C. & Papazoglou, D. *Nature* **287**, 322–324 (1980).
8. Heiken, G. & McCoy, Jr., F. J. *J. geophys. Res.* **89**, 8441–8462 (1984).
9. Bond, A. & Sparks, R. S. J. *Jour. Geol. Soc. Lond.* **132**, 1–17 (1976).
10. Watkins, N. D. *et al. Nature* **271**, 122–126 (1978).
11. Rapp, Jr., G., Cooke, S. R. B. & Henrickson, E. *Science* **179**, 471–473.
12. Fornaseri, M., Malpieri, L. & Tomolo, L. *Archaeometry* **17**, 112–116 (1975).
13. Van Liere, W. J. *Ber. Ryked. Oudheidk. Bode-monderz* **10/11**, 7–69 (1961).
14. Emery, K. O. & Neev, D. *Bull. Israel Geol. Surv.* **26**, 1–24 (1960).
15. Melson, W. G. & Van Beek, G. *Soc. archeol. Sci. Newsl.* **7**, 3–6 (1983).
16. Lacombe, H. & Tchermia, P. in *The Mediterranean Sea—A Natural Sedimentation Laboratory* (ed. Stanley, D. J.) 25–36 (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1972).
17. Ninkovich, D. & Heezen, B. C. in *Proc. Seventeenth Symp. Colston Res. Soc.*, 413–452 (Butterworths, London, 1965).
18. Ninkovich, D. & Heezen, B. C. *Nature* **213**, 582–584 (1967).
19. Keller, J., Ryan, W. B. F., Ninkovich, D. & Altherr, R. *Bull. Geol. Soc. Am.* **89**, 591–604 (1978).
20. Federman, A. N. & Carey, S. N. *Quat. Res.* **13**, 160–171 (1980).
21. McCoy, F. W. in *Thera and the Aegean World II* (ed. Doumas, C.) 57–78 (Aegean World, London, 1980).
22. Vinci, A. *Mar. Geol.* **64**, 143–155 (1985).
23. Galanopoulos, A. G. *Alterum* **10**, 131–137 (1964).
24. Van Bemmelen, R. W. in *Acta First int. Scient. Congr. on the Volcano of Thera, 1969* (ed. Marinatos, S.) 5–50 (Archaeological Services of Greece, Athens, 1971).
25. Bennett, J. G. *Systematics* **1**, 127–156 (1963).
26. Fisher, R. V., 1964. *J. geophys. Res.* **69**, 341–357 (1964).
27. Shaw, D. M., Watkins, N. D. & Huang, T. C. *J. geophys. Res.* **79**, 3087–3094 (1974).
28. Maldonado, A. & Stanley, D. J. in *Sedimentation in Submarine Canyons, Fans and Trenches* (eds Stanley, D. J. & Kelling, G.) 239–257 (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania 1978).
29. Sparks, R. S. J., Brazier, S., Huang, T. C. & Muerdter, D. *Mar. Geol.* **54**, 131–167 (1984).
30. Stanley, D. J. & Liyanage, A. N. *Mar. Geol.* (in the press).
31. El Askary, M. A., Abdel Aal, A. A. & Frihy, O. E. *N. Jb. Geol. Palaont. Abh.* **167**, 347–374 (1984).
32. Michael, H. N. in *Thera and the Aegean World I* (ed. Doumas, C.) 791–795 (Aegean World, London, 1978).
33. Biddle, M. & Ralph, E. K. in *Thera and the Aegean World II* (ed. Doumas, C.) 247–252 (Aegean World, London, 1980).
34. Downey, W. S. & Tarling, D. H. *Nature* **309**, 519–523 (1984).
35. Vitaliano, D. B. *Legends of the Earth* (Indiana University Press, Bloomington, 1973).
36. Downey, W. S. & Tarling, D. H. *New Scientist* **103**, 49–52 (1984).

Impact of the North African drought and El Niño on mineral dust in the Barbados trade winds

Joseph M. Prospero & Ruby T. Nees

Division of Marine and Atmospheric Chemistry,
Rosenstiel School of Marine and Atmospheric Science,
University of Miami, Miami, Florida 33149, USA

A large area of North Africa has been affected by a drought which began in the late 1960s and which has resulted in widespread crop failures, starvation and death. At the same time, there has been a great increase in the concentration of mineral aerosol in the trade winds over the western North Atlantic Ocean and the Caribbean, especially in the early 1970s and again in the early 1980s, when the drought was most severe. Concentrations in 1983 and 1984 were unprecedented and exceeded pre-drought means by a

factor of four. Here we show that spring–summer dust concentrations at Barbados ($13^{\circ}10' \text{ N}$, $50^{\circ}30' \text{ W}$) are correlated to rainfall deficits in the sub-Saharan. Winter dust concentrations, which are normally at a seasonal low, also increased during the drought; concentrations in 1982–1983 were ten times greater than pre-drought values, and appeared to be related to large-scale circulation changes associated with El Niño rather than to the drought itself.

We have measured various properties of aerosols over the equatorial and tropical North Atlantic for almost two decades. We^{1–7} and others^{8–12} have found that one of the principal aerosol components is mineral dust originating from the arid and semi-arid regions of West Africa. Satellite imagery shows that these outbreaks often cover $10\text{--}15^{\circ}$ of latitude in the trade-wind belt and that they frequently extend over much of the tropical North Atlantic^{5–7,13}. During intensive field studies, measurements made concurrently at ships and island stations (including Barbados in the West Indies) yielded mean dust concentrations that were relatively uniform over large areas⁴. Information regarding longer-term variation comes from Barbados, where an aerosol study program has been in effect since 1965 at sites on the east coast^{1,8}. Because of the absence of electrical power before 1971, routine collections were initially made by means of nylon meshes suspended in the wind at a site on the east coast of the island⁸; during field experiments, samples were collected using filters and pumps^{1,6,7}. Since 1971, all samples have been collected with filters and pumps sampling at rates of $1\text{--}2 \text{ m}^3 \text{ min}^{-1}$. Sampling has been continuous except during equipment failures and several periods of extended shutdown. The mineral aerosol concentration on the filters is determined by first extracting the water-soluble components and then ashing the filter at 500°C . The residue is assumed to be mineral matter on the basis of supporting studies of elemental and mineralogical composition^{2,5}. The mass collection efficiency of the filters (IPC 1478 and Whatman 41) for soil aerosols is essentially 100%, and that of the meshes (determined by on-site intercomparisons with filters) is 30% (ref. 1). The sampling duration for individual filters ranged from 12 hours to one week, depending on the experimental protocol in effect. Monthly mean mineral aerosol concentration data are summarized in Fig. 1.

Two major features stand out in Fig. 1. First, there is a clearly defined annual cycle in dust transport, with the maximum occurring in the summer months (usually June or July), when means are more than 10 times greater than those of the winter months. Although there are significant year-to-year variations, the period of increased transport usually begins in April or May and ends in September or October. (The winter of 1982–83 is an anomaly and will be discussed later.) Second, there were two periods when dust concentrations increased sharply: one in the early 1970s, as reported previously¹, and the second in the early 1980s. The annual mean concentrations in 1983 and 1984 (18.7 and $16.4 \mu\text{g m}^{-3}$ respectively) were the highest ever recorded; for comparison, the concentrations in 1966, 1967 and 1968 were 3.9 , 4.5 and $3.2 \mu\text{g m}^{-3}$ respectively.

The periods of unusually high dust concentrations in Barbados correspond to times of severe drought in Africa^{14–16}. For reference we use Lamb's^{14,15} rainfall data (from 20 stations that lie between 11° and 18° N and west of 10° E), which are presented in terms of departures (measured in terms of standard deviations from the mean) from a base period, 1941–74. Four distinct epochs are evident in the record: six dry years in the 1940s; abundant rain between 1950 and 1958; normal rain, 1959–67; drought, 1968–84. The duration and intensity of the 1968–84 drought are unprecedented in this century.

Figure 2 shows the correlation between rainfall deficits in the sub-Saharan and dust concentrations on Barbados, the highest dust concentrations being associated with the peaks of the droughts in 1972–73 and 1983–84. However, the data also suggest that factors other than rainfall can be important. For example, rainfall in the mid-1960s was close to normal, but the dust

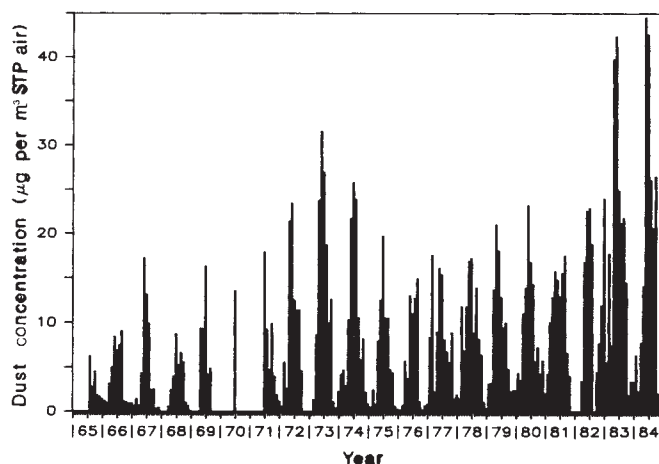


Fig. 1 Monthly mean trade-wind mineral aerosol concentration at Barbados, West Indies from 1965 to 1984. The means for all plotted values after early 1973 are based on at least 21 days of sampling per month; some means prior to that time are based on shorter time periods, none shorter than 10 days¹. Data from 1965 to September 1966 are from ref. 8 and A. C. Delany (personal communication). Before 1971, data were collected by a semi-quantitative technique (see text, refs 1, 8), except during the summer of 1968, when filter samples were collected from aircraft^{6,7}.

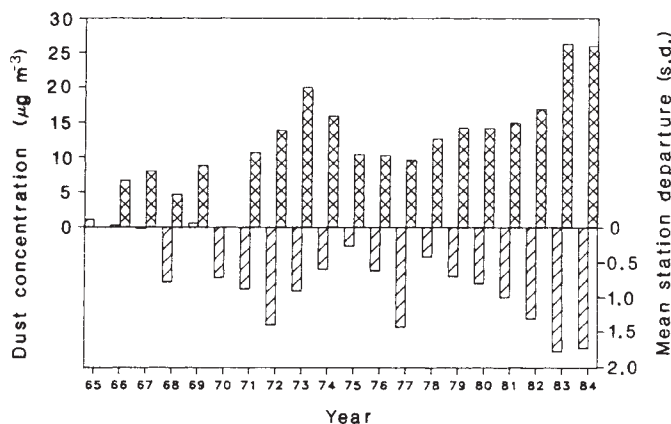


Fig. 2 Mineral aerosol concentrations at Barbados (hatched) and rainfall departures from the mean in the sub-Saharan (diagonally ruled) for the period 1965–84. Dust concentrations are means for the period April to September, which encompasses a major portion of the dust transport period observed at Barbados. Annual rainfall departures are expressed as standard deviations from the mean for the base period 1941–74 (refs 14, 15: P. J. Lamb, personal communication).

concentrations for June and July of 1967 (Fig. 1) were much higher than those for the years 1965, 1966 and 1968, and are comparable to those obtained in the mid-1970s when the region was suffering from years of persistent drought. However, on the basis of the mean for the April–September period (Fig. 2) or on an annual basis the dust concentrations for 1967 are not greatly different from those of adjacent years.

Similarly, it is clear that a short period of drought does not result in an immediate and major increase in dust mobilization and transport. For example, drought began in 1968 with a severe rainfall deficit, yet dust concentrations in the summers of 1968 and 1969 (Fig. 2) were not unusually high. Again in 1977, rainfall deficits were unusually severe; yet dust concentrations in 1977 and the following several years increased by only a relatively small amount.

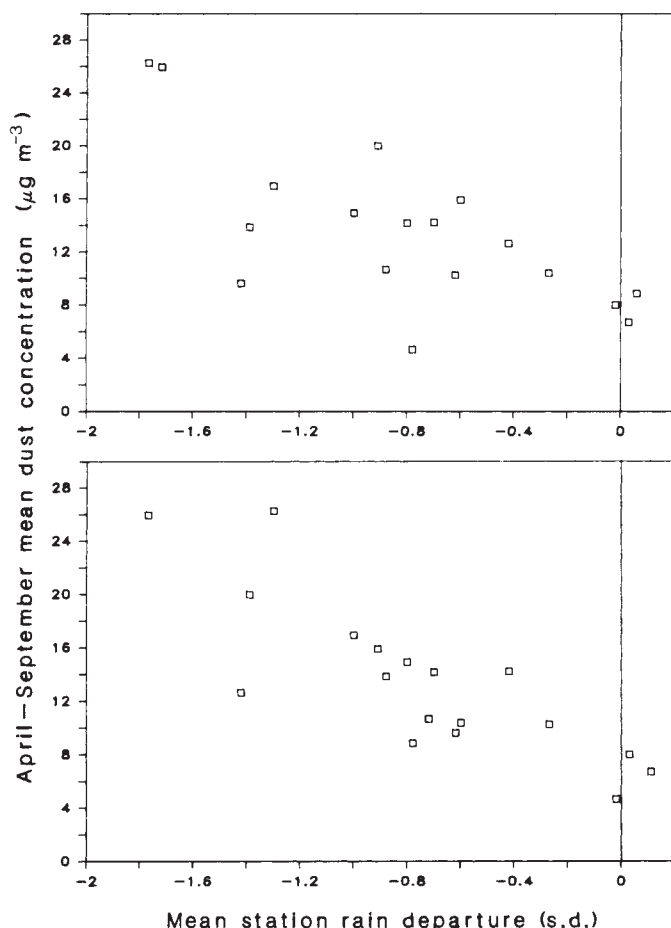


Fig. 3 Mean April-September dust concentration at Barbados plotted against mean station rain departures^{14,15} (expressed as standard deviations from the mean) in the sub-Sahara. *a*, Dust concentration versus current year departure; *b*, dust concentration versus previous year departure.

Scatter diagrams of mean summer dust concentrations versus rainfall departures (Fig. 3*a, b*) suggest that dust transport is better correlated with rainfall data of the previous year than with that of the current year. Similar diagrams (not shown) indicate that the correlation improves if the dust concentration is plotted against the mean of the departures for the previous two or three years. This finding supports the generally held belief that soils become more vulnerable to deflation as drought conditions are prolonged.

Winter dust concentrations are much more variable from year to year than are summer concentrations (Fig. 4); nonetheless, it is clear that the mean winter values have also increased substantially during the drought. In particular, dust concentrations during the winter of 1982-83 were extremely high; the December 1982 and January 1983 means were, respectively, 12 and 25 times those for the period 1965-78. Under normal conditions, winter concentrations over the tropical North Atlantic are quite low and the annual cycle of transport does not begin until February or March³. Because of the large-scale circulation patterns usually in effect, the transport, once it begins, takes place primarily in the latitudes south of Barbados. Thus, early in the year, high concentrations of African dust are observed along the north-east coast of South America, while concentrations remain relatively low at Barbados³.

We attribute the high dust concentrations in the winter of 1982-83 to the anomalous large-scale circulation patterns associated with the recent El Niño/Southern Oscillation (ENSO) event. Global wind field analyses for December, January and

February, 1982-83¹⁷, show that, at 850 mbar (1.5 km altitude), the winds coming off the west coast of Africa were unusually strong and the equatorial trade winds had a much stronger westward component. Such conditions would facilitate transport in the latitude of Barbados. The fact that winter dust concentrations at Barbados are so poorly correlated with precipitation in Africa (not shown) is further evidence that meteorological factors such as large-scale circulation patterns are relatively more important than precipitation in effecting transport at this time of year. Variations in large-scale circulation pattern also appear to be responsible for the unusually high concentrations of dust observed at Barbados in early 1977 and 1978.

Clearly, there has been a dramatic increase in the quantity of dust being transported to the western Atlantic. This increase was noted at least as far north as Miami, Florida, where routine sampling has been carried out since the mid-1970s^{18,19}. Evidence that this increase is chiefly related to processes occurring in North Africa, rather than to changes in transport paths or to removal during transit, comes from an examination of NOAA satellite imagery (M. Matson, personal communication), which indicates that during the past several years, dust outbreaks along the coast of West Africa appear to have been unusually frequent and intense, especially in 1983. Also, meteorological reports show that the frequency of dust storms in Mauritania has increased by at least several hundred per cent during the past few years relative to the pre-drought period²⁰. However, it is impossible to identify the specific sources or processes responsible for this increase. The factors affecting deflation are many and complex, and they are poorly understood even in areas that are relatively accessible and well studied^{21,22}. Nonetheless, our data and other observations suggest that the frequency and intensity of deflation events have increased during the drought and that the increase is related in some way to rainfall deficits, either directly or indirectly.

Our data are unique in that we relate quantitative measures of dust transport to short-term rainfall departures from the climatic mean. In effect, we are assessing the dust-yield response of an area to short-term meteorological and climatic change. Our data differ from previous climatological studies²³ that show a poor correlation between dust storm frequency (as obtained from meteorological reports) and absolute precipitation amounts, with the frequency decreasing for rainfall amounts below 100-200 mm yr⁻¹. Hyperarid regions are good sources of sand, but apparently not of fine dust.

Thus, our results suggest that the dust yield of a region can be an extremely sensitive indicator of the transition from moist to arid climate, although the dust yield could conceivably decrease over long timescales, as a new climatic equilibrium is

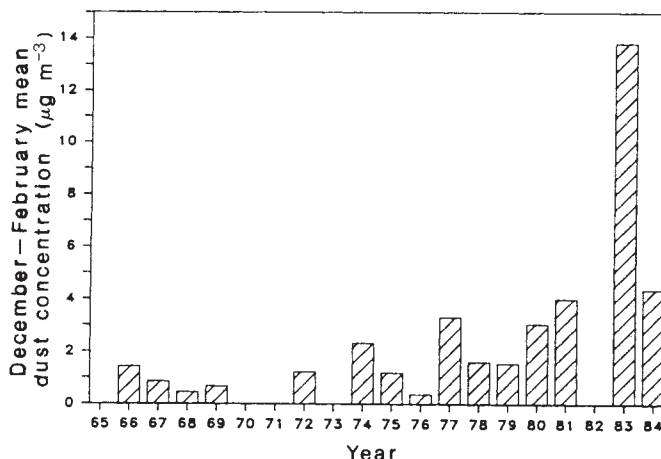


Fig. 4 Mean trade-wind mineral aerosol concentrations for the months December to February at Barbados. The data are plotted for the year in which the January and February data fall.

attained. We conclude that time-series measurements of aeolian materials in deep-sea sediments⁵ are a valid means of assessing the degree of aridity on the continents and of inferring the strength of the large-scale wind systems²⁴; however, the processes of dust generation and transport are complex and depend on many factors, especially on the process of climatic change. Until we understand these processes, interpretations of the palaeoclimatic record on the basis of aeolian components will be somewhat speculative²⁵.

Another relevant factor is the possible impact of dust on climate²⁶. High concentrations of dust can reduce the downward radiative flux to the surface while at the same time increasing the heating of the atmosphere. African dust outbreaks over the Atlantic can produce extinction optical depths of ~ 1 over areas of 10^6 km^2 (refs 10, 11); convection is strongly suppressed under such conditions. In Nigeria, advected dust events produce optical depths up to 2, with an associated reduction in daily mean total solar radiation of 28% and temperature decreases of up to 6°C ²⁷. Optical depths of this magnitude are comparable to those anticipated in a number of 'nuclear winter' scenarios²⁸. Computer climate model simulations²⁶ indicate that high concentrations of dust over Africa could increase atmospheric stability to such a degree as to inhibit convection and decrease the probability of rainfall; such a process could perpetuate drought.

Arid regions comprise about 30% of the continental surface area, and dust from large arid areas such as those in Asia²⁹⁻³¹ could conceivably have a major impact on the chemical and physical properties of the atmosphere. However, the African sources appear to be unique in their strength and persistence and, to our knowledge, the current rate of transport of dust out of North Africa is unprecedented in recent times.

We thank C. Shea for operational assistance on Barbados and E. Manning for use of the sampling site. The work was supported by NSF grants ATM-8016127 and ATM-8209759.

Received 22 July 1985; accepted 21 February 1986.

- Prospero, J. M. & Nees, R. T. *Science* **196**, 1196-1198 (1977).
- Glaccum, R. A. & Prospero, J. M. *Mar. Geol.* **37**, 295-321 (1980).
- Prospero, J. M., Glaccum, R. A. & Nees, R. T. *Nature* **289**, 570-572 (1981).
- Prospero, J. M., Savoie, D. L., Carlson, T. N. & Nees, R. T. in *Saharan Dust: Mobilization, Transport, Deposition* (ed. Morales, C.) 171-186 (Wiley, New York, 1979).
- Prospero, J. M. in *The Sea* Vol. 7 (ed. Emiliani, C.) 801-874 (Wiley, New York, 1981).
- Prospero, J. M. & Carlson, T. N. *J. geophys. Res.* **77**, 5255-5265 (1972).
- Carlson, T. N. & Prospero, J. M. *J. appl. Met.* **11**, 283-297 (1972).
- Delany, A. C. et al. *Geochim. cosmochim. Acta* **31**, 885-909 (1967).
- Schutz, L. *Ann. N.Y. Acad. Sci.* **338**, 515-532 (1980).
- Carlson, T. N. & Benjamin, S. G. *J. atmos. Sci.* **37**, 193-213 (1980).
- Carlson, T. N. & Caverly, R. S. *J. geophys. Res.* **82**, 3141-3152 (1977).
- Carlson, T. N. *Mon. Weath. Rev. U.S. Dep. Agric.* **107**, 322-355 (1979).
- Prospero, J. M. & Carlson, T. N. in *Weather and Weather Maps* (ed. Liljequist, G. H.) 677-691 (Birkhauser, Basel, 1981).
- Lamb, P. J. *J. clim.* **3**, 419-422 (1983).
- Lamb, P. J. *Z. Gletscher. Glazialgeol.* (submitted).
- Nicholson, S. E. *Mon. Weath. Rev. U.S. Dep. Agric.* **111**, 1646-1654 (1983).
- Climate Diagnostics Bull.*, November (NOAA National Weather Service, Washington, 1983).
- Carder, K. L., Steward, R. G., Betzer, P. R., Johnson, D. L. & Prospero, J. M. *J. geophys. Res.* (in the press).
- Prospero, J. M., Nees, R. T. & Uematsu, M. *Eos* **66**, 1309 (1985).
- Middleton, N. J. *Nature* **316**, 431-434 (1985).
- Pewe, T. L. (ed.) *Spec. Pap. geol. Soc. Am.* **186**, (1981).
- Morales, C. (ed.) in *Saharan Dust: Mobilization, Transport, Deposition*, 119-132 (Wiley, New York, 1979).
- Goudie, A. S. *Prog. phys. Geogr.* **7**, 502-530 (1983).
- Rea, D. K., Leinen, M. & Janacek, T. R. *Science* **227**, 721 (1985).
- Prospero, J. M. *Nature* **315**, 279-280 (1985).
- Cookley, J. A. Jr & Cess, R. D. *J. atmos. Sci.* **42**, 1677-1692 (1985).
- Brinkman, A. W. & McGregor, J. Q. *J. R. met. Soc.* **109**, 831-847 (1983).
- National Research Council *The Effects in the Atmosphere of a Major Nuclear Exchange* (National Academy Press, Washington, 1985).
- Uematsu, M. et al. *J. geophys. Res.* **90**, 5343-5352 (1983).
- Blank, M., Leinen, M. & Prospero, J. M. *Nature* **314**, 84-85 (1985).
- Uematsu, M., Duce, R. A. & Prospero, J. M. *J. atmos. Chem.* **3**, 123-138 (1985).

Pliocene change in Pacific-plate motion

Fred F. Pollitz

Department of Geological and Geophysical Sciences,
Princeton University, Princeton, New Jersey 08544, USA

The absolute motion of the Pacific plate has been nearly constant during the past 15 Myr, but Epp¹ recognized that some change took place between 15 Myr and the present, and tentatively chose 1 Myr for the time of the change. Cox and Engebretson² found a change in Pacific absolute motion based on hotspot tracks and a change in Pacific-Antarctic relative motion. The Pliocene deformation in California and the Juan de Fuca rotations led them to postulate a 5-Myr age for the change. The manner and timing of the change are explored more thoroughly here, and we put the time for the change in motion of the Pacific plate at between 5 and 3.2 Myr, basing the younger limit on the correlation of computed hotspot tracks with Pacific and Nazca hotspot traces. The change may explain several observations on the neighbouring continental and oceanic plate boundaries, including direct adjustments at the Pacific spreading ridges. A likely mechanism for the change is a large clockwise torque caused by acceleration of the sinking slab beneath the Kurile trench, where a fossil spreading ridge is inferred to have been subducted.

The 15-~3 Myr segments of every young Pacific chain, with the exception of the Marquesas, are fit very well by Epp's 15-1 Myr stage pole ($65^\circ\text{N } 40^\circ\text{W}$; $0.84^\circ\text{Myr}^{-1}$). There is, however, a large discrepancy between the observed and predicted trends of the Caroline Islands (Fig. 1f) after 5.2 Myr (the age of Ponape Island³) and of the Nazca island chains (Galapagos and Juan Fernandez) between 0 and ≥ 2.5 Myr if we accept Epp's¹ pole and the Nazca-Pacific relative motion of

the past 2.5 Myr (since anomaly 2' time). I therefore regard 2.5 and 5.2 Myr as approximate lower and upper limits to the time of change in Pacific absolute motion. For simplicity, I consider an absolute-motion model characterized by one abrupt change at some time within this interval.

Table 1 lists the migration rates and trends of hotspot volcanic chains. The Pacific island and seamount chains are all well dated⁷, and all contain islands or seamounts (or volcanos) younger than 3 Myr. The age/distance slopes of McDougall and Duncan⁵ are particularly applicable to the young ends of hotspot island chains. The trends are not highly dependent on the time of the change within the given limits. Every chain in Fig. 1, with the exception of the Guadelupe, contributes a rate and or a trend to Table 1, and the migration rate of the Marquesas Islands is also included. I have doubled McDougall and Duncan's⁵ estimate of the standard deviation for the migration rate of the Kodiak-Bowie chain because of its proximity to the pole. The relative motion of the Pacific and Nazca plates was held fixed according to the RM2 model of Minster and Jordan⁸, and absolute motions were obtained by the inversion technique of Minster et al.⁹.

The good fit of the migration rates and trends of the Pacific and Nazca island chains in Table 1 strongly suggests that the absolute motion of the Pacific plate changed from ($65^\circ\text{N } 40^\circ\text{W}$; $0.84^\circ\text{Myr}^{-1}$) to ($61^\circ\text{N } 95^\circ\text{W}$; $0.99^\circ\text{Myr}^{-1}$) (Fig. 2) between 2.5 and 5.2 Myr. Hotspot tracks for the seven island chains of Fig. 1 were generated from the above poles with abrupt changes at several trial times. The youngest transition time for which a good fit is found for all the chains is 3.2 Myr. It is therefore regarded as the upper limit for the time of the change in plate motion, and Epp's² pole is now regarded as a 15-3.2 Myr stage pole. The change was clearly large, as the 95% confidence ellipses for the 15-3.2 Myr and 3.2-0 Myr stage poles (Fig. 2) are significantly different.

Figure 1 shows the changes in trend of the young Pacific island chains required by the suggested change in plate motion.

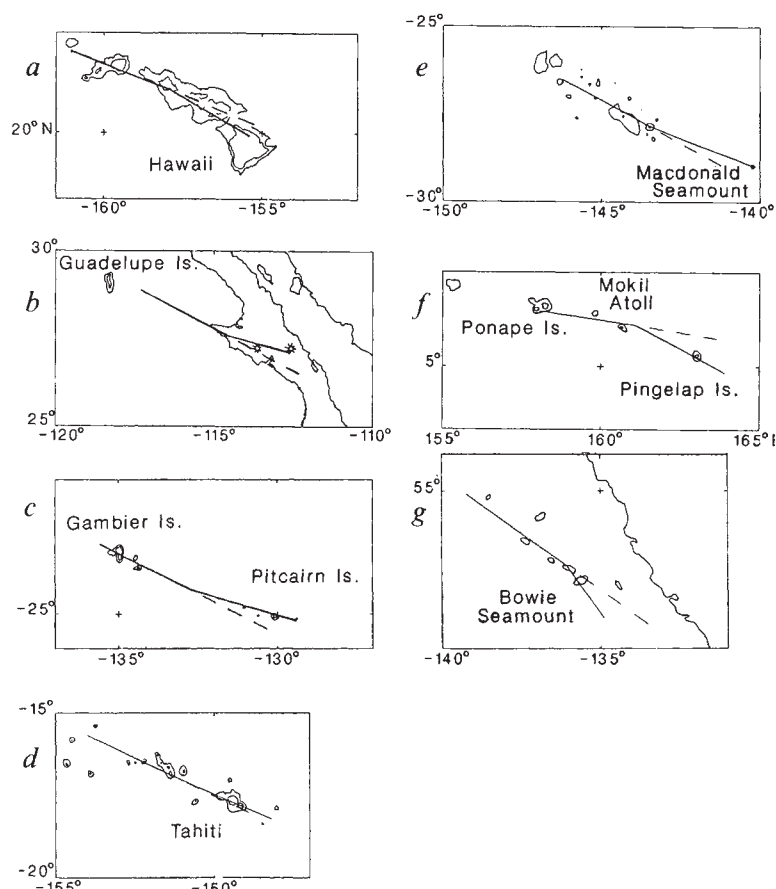


Fig. 1 Hotspot tracks computed for the *a*, Hawaiian; *b*, Guadelupe; *c*, Pitcairn-Gambier; *d*, Society; *e*, Austral; *f*, Caroline; and *g*, Kodiak-Bowie chains, shown by the dark lines. The dashed lines indicate the expected continuation of each chain if there had been no change in plate motion at 3.2 Myr. The tracks are for the time interval 0-7 Myr.

The dashed lines indicate the expected continuations of the island chains if there had been no change at 3.2 Myr. The actual continuations of the Caroline, Austral and Kodiak-Bowie chains clearly diverge from these lines, and the Hawaiian chain bends 8° towards the south. The Society chain is hardly deflected, but its migration rate of volcanism is predicted to have increased by 25 mm yr^{-1} , although the age data is inadequate to verify this. The Pitcairn-Gambier chain, particularly the trend of Pitcairn and its neighbouring seamounts, is fit very well. The Guadelupe chain is revived by a chain of volcanos in Baja California, terminating at the presently active Tres Virgenes volcano.

The convex-northward configuration of the Ponape-Mokil-Pingelap chain in the Caroline Islands suggests that the change in plate motion may have been a gradual process beginning several hundred thousand years before 3.2 Myr, or perhaps as early as 5 Myr (ref. 2). Figure 3 shows the hotspot track in the Caroline Islands resulting from a gradual change in motion between 3.2 and 5.0 Myr. It is a noticeable improvement over Fig. 1*f*, and the Hawaiian and Guadelupe chains are also fit better by a gradual change. Hence, the hotspot record points towards a gradual rather than an abrupt change.

The arrows in Fig. 2 show the difference between the Pacific absolute motion before and after the change. Each velocity

Table 1 Migration rates and trends of young hotspot traces

Pacific hotspots	Lat. (°N)	Long. (°W)	Rate (cm yr ⁻¹)	s.d.	Model*	Ref.
Hawaii	21	157	9.41	0.25	9.30	5
Society	-17	150	10.62	0.97	10.95	"
Marquesas	-10	140	10.13	1.75	10.75	"
Austral	-29	140	10.42	1.56	10.88	"
Kodiak-Bowie	53	137	4.19	0.40	4.40	"
Trend (deg.)†						
Hawaii	21	157	305	10.0	301	This work
Austral	-29	140	292	10.0	290	"
Society	-17	150	299	10.0	294	"
Caroline	7	200	296	10.0	298	"
Pitcairn-Gambier	-25	130	280	10.0	286	"
Nazca hotspots						
Galapagos	-1	91	098	14.0	101	6
Juan Fernandez	-34	81	086	15.0	092	This work

* 3.2-0 Myr stage poles 60.60° N ; 95.10° W ; $0.985^\circ \text{ Myr}^{-1}$; † Due north is 360° . Rates have been adjusted to the timescale of Mankinen and Dalrymple⁴.

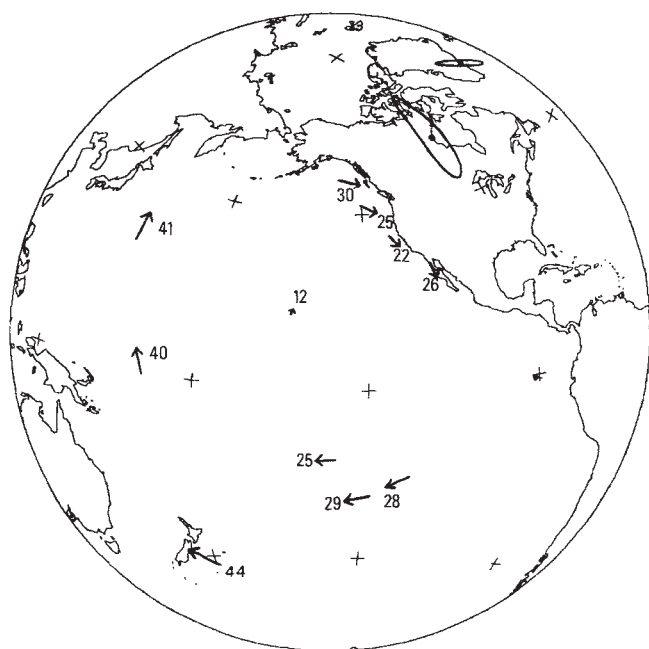


Fig. 2 Differential Pacific motion (that is, the change in motion at 3.2 Myr) for the hotspots and continental plate boundaries discussed in the text and for the northwestern Pacific. The numbers refer to rates in mm yr^{-1} . Each velocity vector lies on a small circle about a pole at $14^\circ \text{ N } 141^\circ \text{ W}$. The 95% confidence ellipses about the 3.2-0 Myr stage pole ($61^\circ \text{ N } 95^\circ \text{ W}$) and 15-3.2 Myr stage pole ($65^\circ \text{ N } 40^\circ \text{ W}$) (ref. 1) are also shown.

vector of differential motion is parallel to a small circle about $14^\circ \text{ N}, 141^\circ \text{ W}$. If the neighbouring plates did not change their absolute motion at the same time as the Pacific, then the Pacific differential motion at those plate boundaries also represents differential relative motions, the trends of which also lie parallel to small circles about $14^\circ \text{ N}, 141^\circ \text{ W}$.

Figure 2 includes the differential motion on the North American plate boundary; it suggests that the sliprate along the San Andreas fault decreased by about 22 mm yr^{-1} . The Pacific-North America relative motion before the change can be found from the present motion⁸ and the change in Pacific absolute motion. It is generally consistent with geological reconstructions of the California continental borderland and Baja California^{11,12} which call for a Baja trans-peninsular fault. Spreading in the Gulf of California began at about 3.5 Myr (ref. 12). When the Gulf is closed according to the plate model presented here, the change in Pacific motion leads to convergence at a rate of $2-4 \text{ mm yr}^{-1}$ in the area of the present Coast Ranges, California, and a large divergence at the mouth of the Gulf of California. As a result, the uplift in the Coast Ranges began in the early Pliocene¹ and appears to have reached maximum intensity in the middle

Pliocene¹⁰, and the Baja trans-peninsular fault was abandoned and modern Gulf spreading was rapidly initiated. A more northerly trend ($10-15^\circ$) predicted for pre-Gulf transform faults is verified by the trend of the Tres Marias relative to the Tamayo fracture zone¹².

Major changes in the Juan de Fuca (JDF) spreading centre occurred between 5 and 3 Myr with large clockwise rotations of spreading axes and fracture zones as well as an overall decrease in spreading rate^{13,14}. The decrease in spreading rate is correctly predicted from the Pacific differential motion. Without a change in JDF motion, however, a small counterclockwise rotation of the spreading ridges is predicted. If the pole of differential motion were located a little farther east ($14^\circ \text{ N } 131^\circ \text{ W}$) or the JDF plate motion swung to a more easterly direction, then both the rate decrease and spreading centre rotations could be correctly predicted. The first possibility is within the error limits of the change, and the second possibility could arise from Pacific-JDF coupling through the Blanco and Mendocino fracture zones.

The change in convergence rate from 0 to 44 mm yr^{-1} along the Indian-Pacific boundary at New Zealand (Fig. 2) accounts for the conversion of that plate boundary from a strike-slip to an oblique-slip fault within the past 5 Myr. It is likely that the time of this conversion was early-mid-Pliocene, given the observation¹⁵ that schist fragments from the Southern (New Zealand) Alps are not found in gravels older than the early Pleistocene.

The increased southerly Pacific motion at the Pacific-Rivera (PR) and Pacific-Cocos (PC) plate boundaries is predicted to have led to counterclockwise rotations of their spreading ridges. Such a rotation occurred on the PR spreading centre at 3.5 Myr (ref. 12), and a similar rotation is observed on the PC spreading centre between ~ 6.5 and 3.5 Myr (refs 16, 17). As shown in Fig. 4, the Nazca-Pacific differential motion provides an explanation for the onset of asymmetric accretion on the East Pacific Rise, south of the Garret fracture zone, at 2.5 Myr or earlier¹⁸, with the higher spreading rates on the Nazca side. The shift in Pacific absolute motion required little change in the Nazca component of spreading north of the Garret fracture zone, but substantial increases south of Garret. Between 21° and 23° S , the shift required very large increases in spreading rate in a short time. This could explain the origin of the Easter plate at 3.2 Myr (ref. 19). The transverse component of differential motion against the Nazca plate may have caused a counterclockwise rotation of the spreading axis north of the Garret fracture zone, leading to the formation of the distinct, right-stepping transform domains recognized throughout that area²⁰.

The Pacific stage poles of Cox and Engebretson² require the pole of differential motion to be located at $38^\circ \text{ N } 92^\circ \text{ W}$. While this pole predicts the proper amount of convergence in California and New Zealand, it seriously mispredicts the continuations of the Austral, Pitcairn-Gambier, and Guadelupe chains. The correct prediction of these chain trends and of the tectonic evolution of the northern East Pacific Rise, requires greater

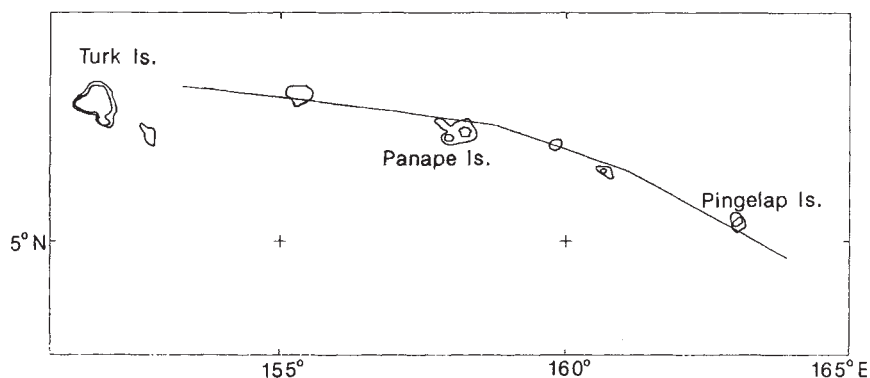


Fig. 3 The hotspot track for the Caroline chain, computed assuming a gradual change between 3.2 and 5.0 Myr. The track corresponds to the time interval 12.8 Myr, the age of the oldest rocks found on Truk³.

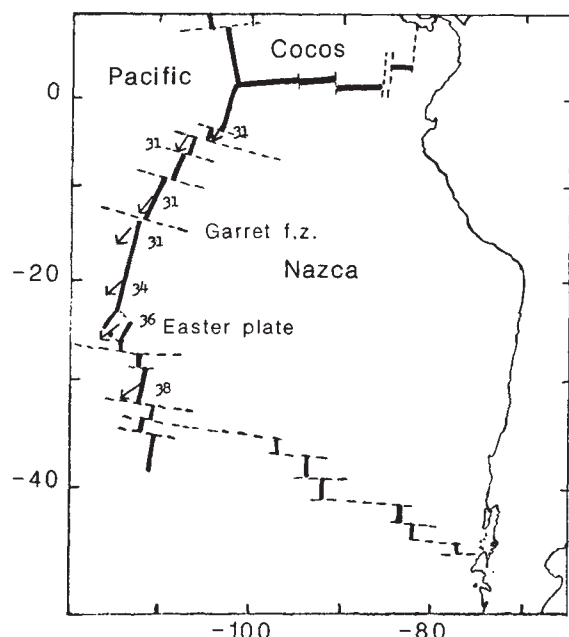


Fig. 4 Differential Pacific-Nazca relative motion is shown by the direction of the arrows. Numbers refer to rates in mm yr^{-1} .

southward movement of the Pacific plate in the entire east-Pacific area, rather than the northward movement predicted by Cox and Engebretson's² model. However, their 5–0 Myr stage pole correctly predicts the post-5-Myr trend of the Marquesas Islands, which is a large misfit in my model. I must assume that the Marquesas hotspot has been migrating rapidly westward for unknown reasons.

The pole of differential motion (Fig. 2) appears to be located near the centre of mass of the Pacific plate. The largest angular accelerations of a spherical plate about its centre of mass result from a torque exerted at positions close to 90° from the centre of mass. The Kuril-Kamchatka trench is about 65° from the centre of mass of the Pacific plate, and is the trench farthest removed from the pole. A large torque in the sense necessary to produce the change would result if the subducting slab began to sink faster. Creager and Jordan²¹ presented evidence that the subducting slab beneath the Sea of Okhotsk extends to a depth of 900–1,000 km, well below the 670-km discontinuity. An average age increase of 50 Myr in 1,000 km of old Kurile slab would increase its negative buoyancy by 10–20%, the correct order of magnitude to produce the change. An ageing of the slab would result if either a fossil spreading ridge or a fossil escarpment (such as the Bauer escarpment²²) were subducted. The age of the oceanic lithosphere now being subducted beneath Japan is about 120 Myr, and the age increases away from the trench, suggesting that a fossil Pacific-Izanagi^{23,24} spreading ridge was subducted perhaps 15 Myr ago. After subduction of the fossil ridge a new trend would have been established, in which progressively older oceanic crust was being subducted, leading to an effective ageing of the slab of tens of millions of years. A change in stress pattern in Japan in the latest Miocene²⁵ introduced a large component of compression normal to the trench which may indicate a large increase of stress in the north-west Pacific as the slab accelerated.

I thank Jason Morgan for a review of the manuscript and assistance with the plotting programs. Marcia McNutt and John Suppe for discussions. Bernard Minster for providing his plate program, and Crystel Bottcher for aid with the figures. This work was supported by NSF grant EAR 83-20104.

3. Keating, B. H. *et al.* *J. geophys. Res.* **89**, 9937–9948 (1984).
4. Mankinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* **84**, 615–626 (1979).
5. McDougall, I. & Duncan, R. A. *Tectonophysics* **63**, 275–295 (1980).
6. Hey, R. N. thesis, Princeton Univ. (1974).
7. Jarrard, R. D. & Clague, D. A. *Rev. Geophys. Space Phys.* **15**, 57–76 (1977).
8. Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331–5354 (1978).
9. Minster, J. B., Jordan, T. H., Molnar, P. & Haines, E. *Geophys. J. R. astr. Soc.* **37**, 541–576 (1974).
10. Christensen, M. N. *Bull. Geol. Soc. Am.* **76**, 1105–1124 (1965).
11. Crouch, J. K. *Bull. Geol. Soc. Am.* **90**, 338–345 (1979).
12. Mammaerickx, J. *Mar. geophys. Res.* **4**, 305–318 (1980).
13. Wilson, D. S., Hey, R. N. & Nishimura, C. *J. geophys. Res.* **89**, 9215–9225 (1984).
14. Riddihough, R. *J. geophys. Res.* **89**, 6980–6994 (1984).
15. Walcott, R. I. *Geophys. J. R. astr. Soc.* **52**, 137–164 (1978).
16. Klitgord, K. D. & Mammaerickx, J. *J. geophys. Res.* **87**, 6925–6950 (1982).
17. Mammaerickx, J. & Klitgord, K. D. *J. geophys. Res.* **87**, 6751–6759 (1982).
18. Rea, D. K. *Geol. Soc. Am. Bull.* **89**, 836–844 (1978).
19. Handschumacher, D. W., Pilger, R. H. Jr, Foreman, J. A. & Campbell, J. F. *Mem. geol. Soc. Am.* **154**, 63–76 (1981).
20. Fox, P. & Gallo, D. G. *Tectonophysics* **104**, 205–242 (1984).
21. Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **96**, 3031–3049 (1984).
22. Rea, D. K. *Geophys. Res. Lett.* **5**, 561–564 (1978).
23. Woods, M. T. & Davies, G. F. *Earth planet. Sci. Lett.* **58**, 161–166 (1982).
24. Engebretson, D. C., Cox, A. & Gordon, R. G. *J. geophys. Res.* **89**, 10291–10310 (1984).
25. Takeuchi, A. *Tectonophysics* (in the press).

Oxygen microprofiles measured *in situ* in deep ocean sediments

Clare E. Reimers, Kathleen M. Fischer, Ray Merewether, K. L. Smith Jr & Richard A. Jahnke

Scripps Institution of Oceanography, La Jolla, California 92093, USA

Diffusive fluxes of oxygen into sediments cannot be calculated reliably without accurate fine-scale concentration measurements very near an undisturbed sediment–water interface. For deep-sea sediments this is difficult, and thus few data on seafloor oxygen gradients have been reported^{1–4}. We present here bottom-water/pore-water oxygen microprofiles that were measured *in situ* with microelectrodes and a free-vehicle microprofiling instrument at ocean depths of 1,275 and 3,790 m. The vertical distribution of oxygen in the upper few centimetres of sediment reflects variations in bottom-water oxygen concentration, oxygen demand and biologically produced small-scale (millimetre to centimetres) sediment heterogeneity and microtopography. Oxygen microprofiles measured on board ship in box-cored sediments from the same sites have features that are similar to the profiles measured *in situ* except within the first 2–4 mm below the sediment surface. This difference is probably caused by changes in boundary conditions and disturbance of the uppermost sediment during coring and sample retrieval.

Microelectrodes were first introduced as tools for obtaining small-scale oxygen distributions across the marine sediment–water interface in shallow waters by Revsbech *et al.*⁵. When used with a separate reference electrode, these probes have a spatial resolution of 10–100 μm , a 90% response time <0.2 s, and are rugged enough to be introduced several centimetres below the sediment–water interface^{3,6,7}. In laboratory tests and actual measurements we have found that the cathode-type oxygen microelectrodes are easily pressure-compensated by filling with a non-conducting fluorocarbon fluid. With increasing pressure, the electrode signal increases approximately exponentially according to thermodynamic principles⁸. At a constant ocean depth the response is proportional to the oxygen partial pressure in the surrounding sea or pore water.

In this study, distributions of dissolved oxygen in surface sediments from two eastern North Pacific stations were investigated with microelectrodes mounted on a self-contained profiling and recording instrument. This instrument is designed to lower (or raise) an 80-cm² circular array of up to seven microelectrodes and one common reference electrode in

Received 14 August 1985; accepted 18 February 1986.

1. Epp, D. thesis Univ. Hawaii (1978).

2. Cox, A. & Engebretson, D. *Nature* **313**, 472–474 (1985).

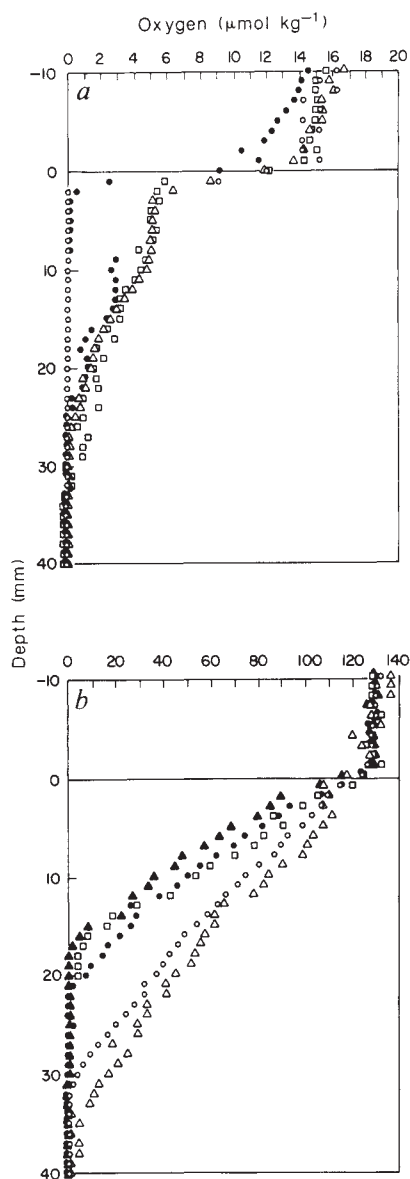


Fig. 1 Oxygen concentrations measured *in situ* as a function of depth in the sediment. *a*, Santa Catalina Basin; *b*, Patton Escarpment. Each symbol designates a separate microelectrode record that is also listed by letter in Table 1 ($\square$, M; $\triangle$, N; $\bullet$, P; $\blacktriangle$, R; $\circ$, S).

programmable vertical steps. The package was positioned inside the base of a tripod frame carrying a weight release system and a flotation assembly.

The stations were sampled in late July 1985, and were located in the Santa Catalina Basin ($33^{\circ}14.8' \text{ N}$, $118^{\circ}33.8' \text{ W}$; depth 1,275 m) and off the Patton Escarpment ($32^{\circ}35.7' \text{ N}$, $120^{\circ}32.1' \text{ W}$; depth 3,790 m). Five individual microelectrodes, spaced 4 cm apart around the perimeter of a circle, were used during each deployment. Measurements were taken by each electrode, both above and below the sediment-water interface, in descending steps of 1 mm, over a total distance of 11 cm. The sensors remained stationary at discrete depth horizons for approximately 45 s. The recorded signal levels ranged from 3 to 600 counts after amplification and analogue-to-digital conversion. These signal distributions were later converted to units of oxygen concentration, and portions of these profiles are shown in Fig. 1.

The depth scale in Fig. 1 was constructed by defining the sediment-water interface as the first point in the profile with an electrode output 2 s.d. less than the mean of the preceding

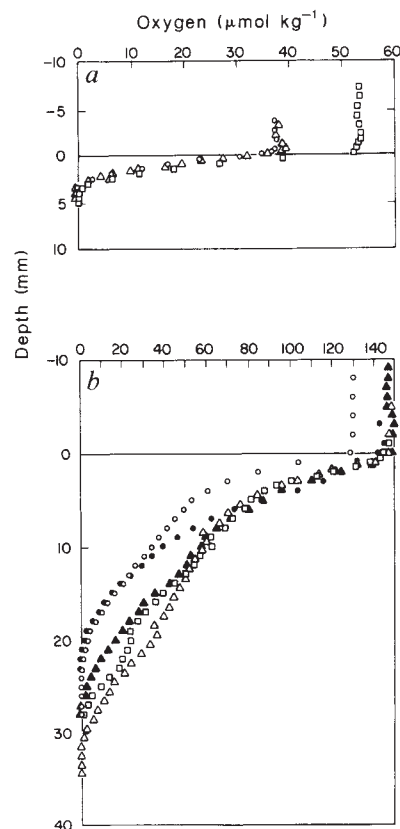


Fig. 2 Oxygen microprofiles measured on board ship in sediment cores. *a*, Three profiles from a box core collected from the Santa Catalina Basin in July 1985; *b*, selected profiles from five separate cores collected off the Patton Escarpment in January 1985 ($\square$, $\triangle$), April 1985 ($\blacktriangle$, $\bullet$) and July 1985 ($\circ$).

15 readings. This means that we have defined the sediment-water interface as an oxygen isopleth rather than as a flat horizon or physical boundary⁹. The oxygen concentration scale was calibrated by equating the bottom-water oxygen concentration at each site to the mean of the (approximately constant) non-zero readings of the part of each profile above the sediment surface. These computations were made after subtracting the residual zero-oxygen current, as determined from the very constant, low-electrode signals at depth in the anoxic layers of the sediments⁶. From the readings above the interface, signal-to-noise ratios and drift rates could also be determined. Both were small, and are listed with other profile characteristics in Table 1. Bottom-water samples were collected by hydrocast in 2-litre Nansen bottles and analysed by a Winkler technique¹⁰. The precision of these O_2 analyses was about $\pm 2 \mu\text{mol kg}^{-1}$.

At the two stations studied, the oxygen concentration gradients near the sediment surface are strongly negative and generally station-specific. However, below the interface, profiles show irregularities and do not reach zero at any consistent depth. We believe such patterns are not due to variable response characteristics of individual electrodes. This conclusion is based on similarities between the microelectrodes in behaviour in laboratory tests, signal-to-noise ratios and drift (Table 1). A more likely explanation for the differences between intra-site profiles is that small-scale (millimetre to centimetres) horizontal concentration gradients arising from sediment heterogeneities or 'microenvironments' were sampled beneath the surface of these deep-marine sediments. These conditions would result if biological activities, such as tube construction, burrowing, feeding and defaecation, caused the sediments to have variable physical properties, or if localized variations exist in oxidation rates¹¹.

Table 1 Microelectrode and oxygen profile characteristics during the *in situ* experiments

Station	Microelectrode	s.d./ $\bar{x}$ *	Drift rate (% mm ⁻¹)	Microtopography† (mm)	ΔO_2 (0–1 mm) ($\mu\text{mol kg}^{-1} \text{mm}^{-1}$)	Cumulative O_2 flux† ($\mu\text{mol per cm}^2 \text{ per day}$)
Santa Catalina Basin	M	0.03	–0.6	+6	6.2	0.07
	N	0.06	–0.9	+5	3.3	0.04
	P	0.03‡	–0.2	+5	6.7	0.07
	S	0.07	0.0	–5	3.1	0.03
					mean $\pm$ s.d.	0.05 $\pm$ 0.02
Patton Escarpment	M	0.03	0.0	+3	14.5	0.15
	N	0.04	–0.1	–2	10.3	0.13
	P	0.01	0.0	+1	10.9	0.11
	R	0.01	0.0	0	9.8	0.14
	S	0.01	–0.3	–3	8.3	0.08
					mean $\pm$ s.d.	0.12 $\pm$ 0.03

Microtopography was estimated by comparing the relative distances the microelectrodes travelled before reaching a gradient, taking into account differences in the lengths of the microelectrodes.

* Standard deviation in microelectrode signal divided by mean signal for last 15 readings before the interface.

† Sign convention as in Figs 1, 2; that is, positive is into the sediments.

‡ Determined above 10 mm altitude.

Although the inference that benthic biological activity influenced our remotely sampled oxygen distributions is speculative, the floor of one of the sites, the Santa Catalina Basin, has been photographed and independently described as being 'littered with maldanid polychaete, foraminiferan tubes, ophiuroid sitzmarks, small burrow openings, and occasional faecal castings'¹². In laboratory studies, Jørgensen and Revsbech⁹ have demonstrated that polychaete tubes and burrows can cause quite dramatic horizontal microelectrode oxygen gradients, and Aller¹³ has modelled the effects of such biogenic structures on variations of other pore-water solutes. Jørgensen and Revsbech⁹ also noted that surface protrusions greater than 0.4 mm in size caused measurable changes in the slopes and spacings of O_2 isopleths surrounding such protrusions. To evaluate the scale of surface microtopography in the vicinity of our one-dimensional profiles, we calculated the differential heights between the profiles at each site (Table 1). The result was that as much as 11 mm of surface topography was sampled at the Santa Catalina Basin site and 6 mm was sampled at the Patton Escarpment. This result implies that a microelectrode could easily have sampled at an oblique angle to the sediment surface, thereby generating an apparently deviant oxygen profile. This 'non-ideal' case is well illustrated in ref. 9. An important consequence of small-scale roughness and heterogeneous sediment properties in deep-sea sediments is that these features should cause rates of solute transfer across the sediment–water interface also to be spatially uneven.

Table 1 lists the approximate magnitudes of, and variations in, oxygen uptake rates calculated from the bottom/pore-water oxygen concentration microprofiles using an integrated version of Fick's first law of diffusion. The expression for these diffusive fluxes is

$$F_{O_2} = \int_0^{x_i} R = \sum_0^{x_i} \left(\left[\phi D_s \frac{dO_2}{dx} \right]_{x_i} - \left[\phi D_s \frac{dO_2}{dx} \right]_{x_{i-1}} \right) \quad (1)$$

where F_{O_2} is a cumulative oxygen diffusive flux, R represents all reactions consuming oxygen, $x_i - x_{i-1}$ are a series of finite depth elements within which net oxygen uptake rates are calculated, and for each depth, ϕ is the porosity, O_2 is the oxygen concentration and D_s is the effective molecular diffusion coefficient. Inherent in these calculations are the assumptions that, when sampled, the pore-water concentrations were in a balance between diffusion and reaction rates (that is, pore-water distributions were at steady state), and that the sediment was sufficiently homogeneous to allow diffusion properties to be

approximated by the relation, $D_s = \phi^2 D_0$, where D_0 is a molecular diffusion coefficient and the porosity ϕ is determined from the water contents of 1–3-mm intervals of sediment cores from each site¹⁴. Derivation of a more realistic diffusive flux model from the fine-scale oxygen distributions would require presently unavailable information on the detail of microtopography and multi-dimensional diffusion patterns around each profile¹³, the relationship of the physical sediment boundary to each measured gradient^{9,15}, and the frequency and magnitude of concentration fluctuations about mean gradients. (At the interface, concentration fluctuations caused by turbulence lead to what is often called the Reynolds flux¹⁶.)

Our goal in future field work will be to improve experimental techniques so that the factors needed to constrain diffusive flux models can be determined. Until then, a conditional evaluation of the relative importance of diffusive exchange and other, non-diffusive, mass transfer processes may be made by comparing total sediment-community oxygen consumption rates, measured by chambers, to the diffusive predictions in Table 1^{14,17}. The total sediment-community oxygen consumption rates measured *in situ*, periodically over 5 yrs, at the basin and escarpment stations^{18,19} are in the range 0.24–0.37 $\mu\text{mol per cm}^2 \text{ per day}$ (21 replicates) and 0.14–0.36 $\mu\text{mol per cm}^2 \text{ per day}$ (19 replicates), respectively. These rates are roughly six and two times the average diffusive estimates, but a factor of two uncertainty for each comparison is likely. Therefore, non-diffusive oxygen transport to the sediment can be positively demonstrated only at the Santa Catalina Basin site. We expect this added flux is the result of the bioturbation processes referred to above.

It is possible that our microelectrode profiles, being once-only, very small-scale measurements, may not represent 'average' regional profiles, and hence sediment–water diffusive exchange, at greater spatial and temporal scales. Some assurance that this is not the case comes from several shipboard oxygen microprofiles measured in cores taken from the Santa Catalina Basin and Patton Escarpment stations (Fig. 2). For this study, our method of measuring oxygen in sediment subcores of larger box cores was altered from that described in the literature^{3,14}, in order to minimize atmospheric contamination of overlying waters. The new method involved submerging the sediments in a temperature-controlled (0–3 °C) seawater bath, bubbled with a N_2 – O_2 mixture which approximated the bottom-water oxygen concentrations.

Comparison of the data in Figs 1 and 2 suggests that the shapes and penetration depths of oxygen profiles measured in

cores on board ship are similar to *in situ* distributions, except in the upper 2–4 mm of the sediment column. Within this depth zone uncharacteristic profiles seem to result from re-equilibration, which occurs because cores are sampled under unnatural boundary conditions (the natural diffusive sublayer is not re-established), or because they have had a history of being jostled, decompressed, warmed and aerated. For example, Fig. 2a shows three vertical microprofiles from a box core taken in July 1985. These profiles display little variability, suggesting that natural heterogeneity was lost due to coring operations, subsampling and re-equilibration within the bath. Alternatively, we may have biased our sampling to 'undisturbed' areas. Note also that the oxygen concentrations in the water above the sediment were higher than bottom-water values and varied during profiling. These problems arose because we had difficulty equilibrating the bath at oxygen concentrations low enough to mimic bottom-water conditions.

Figure 2b shows single profiles from five cores taken at the Patton Escarpment during three cruises in January, April and July 1985. These demonstrate little recognizable spatial or temporal variability and are generally similar to the five *in situ* oxygen profiles from the same site. It is important, however, that the Patton Escarpment core profiles are more strongly concave near the sediment surface than the *in situ* profiles (Fig. 1b). We believe this problem arises even in minimally disturbed cores through some loss or redistribution of surface material. Surface loss artificially shortens and steepens (by re-equilibration) surface gradients, and therefore diffusive flux

rates can be overestimated. Oxygen fluxes into the sediment, estimated from the profiles in Fig. 2b (corrected to true bottom-water concentrations¹⁴) and equation (1), are $0.18 \pm 0.07 \mu\text{mol per cm}^2 \text{ per day}$. These are 1.5 times the mean *in situ* value from the same site.

This research was supported by NSF grants OCE84-15480 and OCE83-15306. We thank G. Gust and T. R. S. Wilson for their critical reviews of early versions of the manuscript.

Received 2 December 1985; accepted 17 February 1986.

- Grundmanis, V. & Murray, J. W. *Geochim. cosmochim. Acta* **46**, 1101–1120 (1982).
- Jahnke, R., Heggie, D., Emerson, S. & Grundmanis, V. *Earth Planet. Sci. Lett.* **61**, 233–256 (1982).
- Reimers, C. E., Kalthorn, S., Emerson, S. R. & Nealson, K. H. *Geochim. cosmochim. Acta* **48**, 903–910 (1984).
- Wilson, T. R. S. *et al. Geochim. cosmochim. Acta* **49**, 811–822 (1985).
- Revsbech, N. P., Jørgensen, B. B. & Blackburn, T. H. *Science* **207**, 1355–1356 (1980).
- Revsbech, N. P. in *Polarographic Oxygen Sensors. Aquatic and Physiological Applications* (eds Nagrauer, E. & Forstner, H.) 265–273 (Springer, New York, 1983).
- Revsbech, N. P. & Jørgensen, B. B. *Adv. Microbial Ecol.* **9**, 293–352 (1986).
- Hitchman, M. L. *Measurement of Dissolved Oxygen* (Wiley, New York, 1978).
- Jørgensen, B. B. & Revsbech, N. P. *Limnol. Oceanogr.* **30**, 11–21 (1985).
- Grasshoff, J. in *Methods of Sea Water Analysis* 59–70 (Verlag Chemie, Weinheim, 1976).
- Jumars, P. A. & Eckman, J. E. in *The Sea* Vol. 8 (ed. Rowe, G. T.) 399–451 (Wiley, New York, 1983).
- Smith, C. R. & Hamilton, S. C. *Deep Sea Res.* **30**, 907–928 (1983).
- Aller, R. C. *Geochim. cosmochim. Acta* **44**, 1955–1966 (1980).
- Reimers, C. E. & Smith, K. L. Jr *Limnol. Oceanogr.* **31**, 305–318 (1986).
- Santschi, P. H., Bower, P., Nyffeler, U. P., Azevedo, A. & Broecker, W. S. *Limnol. Oceanogr.* **28**, 899–912 (1983).
- Boudreau, B. P. & Guinasso, N. L. in *Dynamic Environment of the Ocean Floor* (eds Fanning, K. A. & Manheim, F. T.) 115–145 (Lexington House, Kentucky, 1982).
- McCaffrey, R. J. *et al. Limnol. Oceanogr.* **25**, 31–44 (1980).
- Smith, K. L., Jr, Laver, M. B. & Brown, N. O. *Limnol. Oceanogr.* **28**, 882–898 (1983).
- Smith, K. L. Jr *Limnol. Oceanogr.* (submitted).

Microorganisms in deep-sea hydrothermal plumes

Christopher D. Winn & David M. Karl

Hawaii Institute of Geophysics and Department of Oceanography,
University of Hawaii, Honolulu, Hawaii 96822, USA

Gary J. Massoth

National Oceanic and Atmospheric Administration,
Pacific Marine and Environmental Laboratory,
7600 Sand Point Way NE, Seattle, Washington 98115-0070, USA

The study of hydrothermal vents at oceanic spreading centres has led to a re-evaluation of some basic tenets concerning energy flux through oceanic ecosystems and the sources and sinks for dissolved ions in sea water. Hydrothermal vents vary considerably, from relatively low-temperature (<25 °C) fluid discharges to the spectacular high-temperature (~350 °C) black smokers^{1–3}. The high-temperature vents give rise to buoyant plumes which can be detected hundreds of kilometres away from ridge crests^{4,5}. In the immediate vicinity of hydrothermal vents, chemoautotrophic bacteria are present in vent fluids^{6–8}, attached to rock surfaces^{9,10}, and as endosymbionts in certain macrofauna¹¹. Deep-sea hydrothermal plumes have not, as yet, been examined for the presence of bacteria. We now report the presence of elevated bacterial biomass in a hydrothermal plume originating from the Juan de Fuca Ridge, and discuss the potential geochemical significance of these microorganisms.

During a recent oceanographic expedition (PENTAFLUX 1984), we had an opportunity to sample hydrothermal plumes emanating from the Endeavour and Juan de Fuca Ridges. These oceanic spreading centres are located in the north-east Pacific Ocean, ~450 km west of Seattle, Washington (USA). The DSRV *Alvin* was used to collect water samples in the immediate vicinity of hydrothermal vents on the Endeavour Ridge (47°57' N; 129°04' W), using either acid-cleaned Niskin bottles (2.5 l) or

Table 1 Microbial biomass in hydrothermal fluids

<i>Alvin</i> Dive no.	Sample location	ATP* (ng per l)	Bacteria (cells per l)
1418	Smoker plume (+20 m)	56	†
1418	Smoker plume (+50 m)	14	†
1419	Warm vent	24	1.3×10^8
1419	Warm vent (+6 m)	6.5	7.6×10^7

* Living carbon can be estimated by multiplying ATP by 250 (ref. 14).

† Bacteria were present in these samples but accurate cell counts were unobtainable because of interference from inorganic particulates.

an *in situ* pump. A black-smoker plume was sampled at ~20 and 50 m above a 10-m high chimney, and a warm vent was sampled at and ~6 m above the point of discharge.

Microbial biomass was estimated by measuring the concentration of ATP, and by direct epifluorescence microscopy^{12,13}. ATP concentrations in vent water samples ranged from 6.5 to 56 ng/l, compared with 2 ng/l in ambient bottom water (Table 1). Bacterial cells were present in both the warm-vent and the smoker-plume samples. Quantitative cell counts were not obtained for the smoker-plume samples because of the presence of fluorescent inorganic particles. Magnesium concentrations indicated that all vent water samples were diluted at least 100-fold with ambient sea water.

The temperature responses of the microbial populations in samples collected above the warm vent and above the smoker chimney were determined by measuring the incorporation of ³H-adenine into DNA¹⁵ after a 6-h incubation at one atmosphere and temperatures ranging from 2 to 90 °C (Table 2). Our results demonstrate that the microbial populations from the two vent samples have similar temperature responses, with maximum incorporation at ~30 °C.

A surface ship was used to collect hydrothermal plume samples on the southern Juan de Fuca Ridge (44°38.1' N;

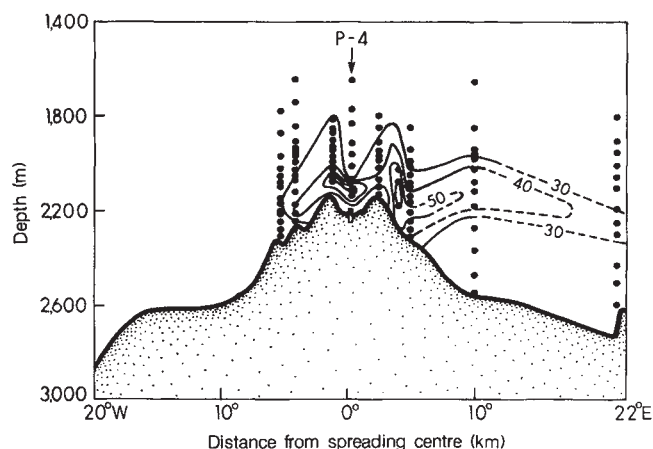


Fig. 1 Contour map of the methane distribution (CH_4 concentrations in nl per l) perpendicular to the crest of the southern section of the Juan de Fuca Ridge. Methane concentrations were determined using a modification of the gas chromatographic method of Swinnerton *et al.*¹⁶. The solid dots show the location of water samples. The data from hydrostation P-4 are shown in Fig. 2.

130°22.9' W). The plume was detected and mapped by an acoustically-tracked sensor package and sampled using CTD/nephelometer rosette hydrocasts⁵. A series of hydrostations was occupied on a transect perpendicular to the ridge crest, which allowed us to contour the distribution of methane (Fig. 1). Water samples were also analysed for particulate ATP, bacterial cell number and total dissolvable Mn (TDMn) at discrete depths from 1200 m to a few metres above the sea floor (Fig. 2). Bacterial cell number and the concentrations of ATP, CH_4 and TDMn were highest at ~100 m above the sea floor. Bacterial cell numbers were low in this profile, relative to other estimates in the Pacific Ocean¹⁷, but the correlation of bacterial cell number and particulate ATP nevertheless suggests that the ATP peak is due to the presence of living bacterial biomass.

We interpret these results as evidence that bacteria growing in environments immediately surrounding high-temperature vents, either in warm vent waters or on the surface of the chimneys, are entrained in the plume as hydrothermal fluids rise and mix with ambient seawater. Although bacteria may grow following entrainment, the biomass observed in the smoker plume can be accounted for by entrainment alone. We found warm vent waters on the Juan de Fuca Ridge to contain up to 500 ng ATP/l (C.W. and D.K., unpublished data), and similar results have been reported from other spreading centres⁷. Moreover, the external surfaces and the porous structure of the

chimneys themselves have been shown to be potential habitats for bacterial growth¹⁰. The similarity of the temperature response of the populations collected at 20 m above the smoker chimney and above the low-temperature vent (Table 2), indicates that the microorganisms within the plume were derived largely from moderate-temperature vent waters. The results of the temperature response experiment do not, however, exclude the possibility that thermophilic bacteria were present in the plume; there was measurable incorporation of ^3H -adenine at the highest temperature tested (90°), indicating the presence of some thermophilic bacteria. These results are consistent with previous measurements in hydrothermal vent samples¹⁸.

At present the geographic extent of microbial enrichment in hydrothermal plumes is unknown. In addition, the data do not allow us to determine whether bacteria continue to grow after entrainment in the plumes. One might question whether bacteria

Table 2 Effects of incubation temperature on the rate of microbial activity

Sample location†	^3H -adenine incorporation into DNA*				
	2 °C	12 °C	30 °C	50 °C	90 °C
Smoker plume (56 ng ATP per l)	4.4	5.2	96.8	16.8	5.0
Warm vent (6.5 ng ATP per l)	1.4	7.2	26.0	4.3	0.7

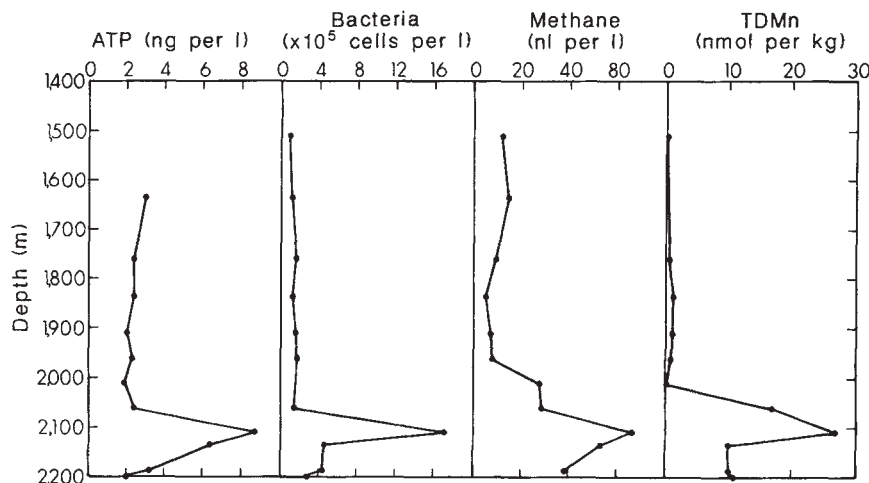
* Incorporation rates in nCi/l after a 6-h incubation at one atmosphere. ^3H -adenine was added at a concentration of 0.1 $\mu\text{Ci}/\text{ml}$ in both experiments.

† Samples were collected 20 m above a smoker chimney and 6 m above a warm vent (see Table 1).

are able to derive sufficient energy for growth in the plume, since potential growth substrates such as CH_4 , H_2 and H_2S are rapidly and substantially diluted upon contact with ambient sea water¹⁹. Our data indicate, however, that a significant microbial biomass does exist in hydrothermal plumes that have been advected away from points of hydrothermal fluid discharge (Fig. 2).

The potential significance of the presence of bacteria within hydrothermal plumes is at least two-fold. If bacteria grow within the water column above vent fields, then these bacteria may be a source of organic carbon and energy for pelagic organisms away from the site of hydrothermal input. Secondly, these bacterioplankton may be important catalysts for the chemical alterations which are known to occur as hydrothermal plumes move off-axis. The residence time of hydrothermally derived H_2 and

Fig. 2 The vertical distribution of ATP concentration, bacterial cell number, methane and total dissolvable manganese (TDMn) at station P-4 on the Juan de Fuca Ridge (see Fig. 1). Total dissolvable manganese concentrations were determined by flameless atomic absorption spectrophotometry.



CH₄ is surprisingly short²⁰, and the chemical mechanism by which these gases are removed from the deep sea is unknown. Since microorganisms can catalyse the oxidation of both CH₄ and H₂, it is reasonable to suggest that microbial activity is responsible for the rapid consumption of these gases. The mechanism by which metals are removed from hydrothermal plumes and deposited in marine sediments is also unknown³. Microbes may play a significant role in this process too, particularly since bacteria encrusted with iron and manganese have been found in the deep sea and have been reported to account for the majority of the vertical flux of these metals^{21,22}.

We thank Kim Kelly-Hansen for methane analyses and Alexander Malahoff and the officers and crew of RV *Discoverer* and RV *Atlantis II* for logistical support. This research was supported, in part, by the NOAA vents program (grant NA80AA-H-00075) and, in part, by NSF (grant OCE83-51751).

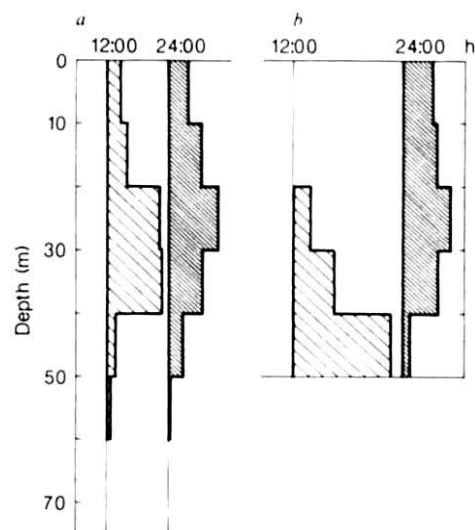


Fig. 1 Noon (12:00) and midnight (24:00) depth distribution (percentage of the total population) of *Cyclops* in August 1985, in Lake Czarny nad Morskim (a) and in Lake Morskie Oko (b). Natural salmonid fish populations have been present in Lake Morskie Oko for millennia, whereas the former lake contains no fish.

be exploited at night, while predators can be avoided during the day by descent to depths where the light intensity is too low for planktivorous fishes. The fitness of a migratory individual should hence be increased if the metabolic cost of migration is lower than the benefit derived from daytime feeding in the upper strata. If this were not the case, other strategies would be selected for, such as that of remaining in the lower strata despite reduced fecundity or in the upper strata despite increased mortality⁷⁻⁹. Migratory behaviour should hence be favoured only in habitats in which both the risk of predation and resource concentration are significantly higher in the upper strata than in the lower strata. Indeed, migratory behaviour does not occur in lakes which show no difference in food concentration between the upper and the lower strata⁹.

Since the end of the last glaciation many alpine lakes above the timberline in Europe and North America have contained no fish. In the Tatra Mountains, a granite formation on the border of Poland and Czechoslovakia, most of these lakes are drained by streams cascading from the moraine edge or flowing underground through thick moraine rubble, both impenetrable to salmonid fishes. Although some of the lakes contain no fish, others were stocked at the end of the nineteenth century and some have been stocked recently (see Table 1). In conditions where the water is very clear, only long-range migrations ensure efficient predator evasion¹⁰, and a prey population can easily be killed off by a predator when food resources are insufficient for energy to be devoted to migration¹¹. The introduction of fish into lakes in the Tatra region led to the extinction of all the cladocerans and most of the copepods^{12,13} that were unable to develop a migratory habit in conditions of low food resources.

The only planktonic crustacean that survived the arrival of fish in the Tatra lakes is the copepod *Cyclops abyssorum*; its survival in lakes stocked with planktivorous brook charr (*Salvelinus fontinalis*) may be related to its reproductive pattern and to the resistance of its eggs to fish digestive enzymes¹⁴. *C. abyssorum* is now present in all the larger Tatra lakes, whether or not they contain fish; it is also the sole planktonic crustacean in many lakes in the Alps which were stocked with fish hundreds of years ago^{12,13}. An unidentified species of *Cyclops* is the sole crustacean survivor in recently stocked lakes above the timberline in the American Rockies¹⁴. In all three locations *Cyclops* is

Received 22 August 1985; accepted 12 February 1986.

1. Corliss, J. B. *et al. Science* **203**, 1073-1083 (1979).
2. Spiess, F. N. *et al. Science* **207**, 1421-1432 (1980).
3. Edmond, J. M., Von Damm, K. L., McDuff, R. E. & Measures, C. I. *Nature* **297**, 187-191 (1982).
4. Craig, H. & Lupton, J. E. in *The Sea* Vol. 7 (ed. Cesare, E.) 391-428 (Wiley, New York, 1981).
5. Baker, E. T., Lavelle, J. W. & Massoth, G. J. *Nature* (in the press).
6. Jannasch, H. W. & Wirsén, C. O. *Bioscience* **29**, 592-598 (1979).
7. Karl, D. M., Wirsén, C. O. & Jannasch, H. W. *Science* **207**, 1345-1347 (1980).
8. Jannasch, H. W. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Bostrom, K., Laubier, L. & Smith, K. L. Jr) 677-709 (Plenum, New York, 1983).
9. Jannasch, H. W. & Wirsén, C. O. *Appl. envir. Microbiol.* **41**, 528-538 (1981).
10. Baross, J. A. & Deming, J. W. *Biol. Soc. Wash. Bull.* **6**, 355-371 (1985).
11. Cavanaugh, C. M., Gardiner, M. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. *Science* **213**, 340-342 (1981).
12. Karl, D. M. & Craven, D. B. *Appl. envir. Microbiol.* **40**, 549-561 (1980).
13. Hobbie, J. E., Daley, R. J. & Jasper, S. *Appl. envir. Microbiol.* **33**, 1225-1228 (1977).
14. Karl, D. M. *Microbiol. Rev.* **44**, 739-796 (1980).
15. Karl, D. M. & Winn, C. D. in *Heterotrophic Activity in the Sea* (eds Hobbie, J. E. & Williams, P. J. LeB.) 197-215 (Plenum, New York, 1984).
16. Swinnerton, J. W., Linnenbom, V. J. & Cheek, H. C. *Analyt. Chem.* **34**, 483-485 (1962).
17. Taga, N. & Matsuda, O. in *Effect of the Ocean Environment on Microbial Activities* (eds Colwell, R. R. & Morita, R. Y.) 443-448 (Univ. Park Press, Baltimore, 1972).
18. Karl, D. M., Burns, D. J., Orrett, K. & Jannasch, H. W. *Mar. Biol. Lett.* **5**, 227-231 (1984).
19. Lupton, J. E., Delaney, J. R., Johnson, H. P. & Tivey, M. K. *Nature* **316**, 621-623 (1985).
20. Welhan, J. A. & Craig, H. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Bostrom, K., Laubier, L. & Smith, K. L. Jr) 391-411 (Plenum, New York, 1983).
21. Cowen, J. P. & Silver, M. W. *Science* **224**, 1340-1342 (1984).
22. Cowen, J. P. & Bruland, K. W. *Deep Sea Res.* **32**, 253-272 (1985).

Predation and the evolution of vertical migration in zooplankton

Maciej Z. Gliwicz

Department of Hydrobiology, University of Warsaw, Nowy Swiat 67, 00-046 Warsaw, Poland

Diel vertical migrations of zooplankton within water bodies have been related to efficient utilization of resources¹⁻⁴ or to avoidance of mortality due to predation^{5,6}. Although the possibility that vertical migrations evolved as an antipredator strategy has attracted much attention, no evidence has yet been presented of predation selecting for traits of migratory behaviour in planktonic animals. Therefore, I have examined nonmigratory and migratory populations of a copepod, *Cyclops abyssorum*, in alpine clear-water lakes in the Tatra Mountains of Poland. I report here that diel vertical migrations are not apparent in lakes without predatory fish, whereas short-range migrations are demonstrable in lakes which have been stocked with planktivorous fishes for decades, and long-range migrations are evident in lakes that have been stocked for centuries or millennia. These observations support the view that migratory behaviour in zooplankton is selected for as a means of evading fish predators.

The predation hypothesis suggests that migrations enable more abundant food resources in the upper strata of water to

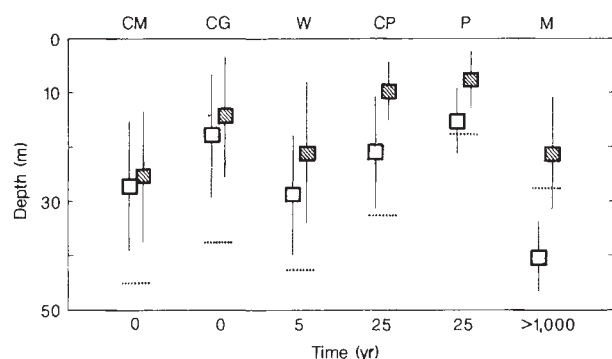


Fig. 2 Noon (□) and midnight (■) mean depth ($\pm$ s.d.) of the *Cyclops* population in August 1985, in six deep lakes: CM and CG (no fish populations); recently stocked lakes W, CP and P; and M (fish present for millennia) (see Table 1 for abbreviations). The abscissa shows the number of years that fish predators have been present in the lakes. The depths receiving 1% of the surface irradiation (Secchi disk transparency multiplied by 2.5) are marked by horizontal dotted lines. Mean counts of *Cyclops* ($\pm$ s.d.) were determined from samples fixed with 4% formaldehyde. Samples were taken in vertical hauls using a quantitative Apstein closing plankton net (0.2 mm mesh size) from discrete 5- or 10-m strata (sample volume of 50 or 100 l, respectively). The total number of *Cyclops* in each set of the noon or midnight samples (from a water column 0.01 m²) ranged from 98 in Lake Czarny nad Morskim to 977 in Lake Wielki.

always accompanied by two or three rotifers of the genera *Polyarthra*, *Keratella* and *Asplanchna*. Rotifers are the main food source for adult *Cyclops*, but scarce phytoplankton composed of small flagellates (*Chrysomonas*) are the food source for nauplii (the earliest larval stage). Rotifers and phytoplankton are most abundant between 5 and 30 m below the water surface¹¹.

The presence of *C. abyssorum* in lakes with different fish-stocking histories provides an excellent opportunity to determine whether vertical migration is related to predation and, if so, how long it took for such a behaviour pattern to evolve.

The depth distributions of *Cyclops* populations at noon and midnight were established from net samples collected between 12 and 18 August 1985—a period of the new moon (dark nights) and cloudless skies (sunny days)—from seven Tatra lakes having different fish-stocking histories. In August, all *Cyclops* populations consisted of juveniles (copepodites). Mature individuals appear only in late winter under thick snow-ice cover, and are soon consumed by fish¹⁴. The migratory behaviour of adults has, therefore, not been confirmed.

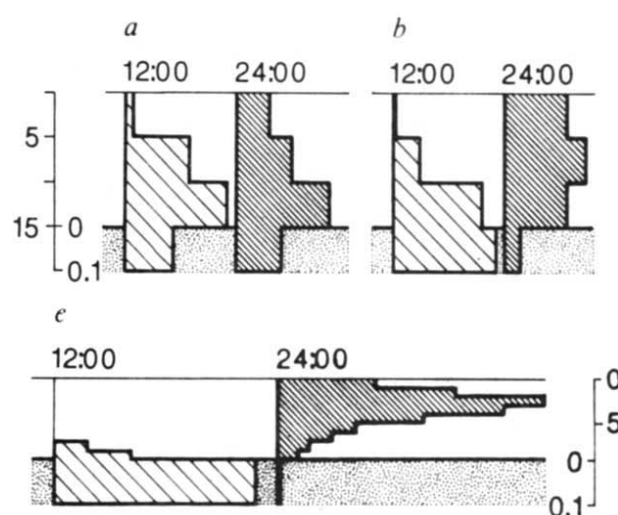


Fig. 3 Summer (August) depth distribution (% of the total population) of *Cyclops* in Lake Zielony stocked for 12 yr (a) and 35 yr (b), and in Lake Gossenköllesee stocked for several hundred years (c). The 1962 distributions in both lakes are based on previous data^{18,19}. In addition to the samples from the water strata, samples from the upper 10 cm of the sediment layer were also included in the estimate of the total population number. The sediment layer is represented by the stippled area.

Figure 1 shows that *Cyclops* does not migrate in Lake Czarny nad Morskim, which contains no fish, but a long-range migration is evident in Lake Morskie Oko, which has contained fish present for millennia. The differences between the mean depths of the *Cyclops* population at noon and midnight (Student's *t*-test) and the variances of the depth distributions (*F*-test) were not significant in Lake Czarny nad Morskim but were highly significant in Lake Morskie Oko (Table 1). The *Cyclops* population in Lake Czarny Gasienicowy is also nonmigratory (Fig. 2); this lake was unsuccessfully stocked several times between 1890 and 1930 and contains no fish. In the three remaining deep lakes, Wielki, Czarny w Pieciu and Przedni, *Cyclops* migration is evident although the range is significantly ($P = 0.002$, *F*-test) lower than in Lake Morskie Oko, which has natural fish populations. In Lake Wielki, which has been stocked for 5 yr, the difference between the noon and midnight mean depths of the *Cyclops* population is significant only at the $P = 0.01$ level, whereas it is significant at $P = 0.001$ in lakes Czarny w Pieciu and Przedni, which have been stocked for 25 yr. The range of migration is

Table 1 Limnology, fish-stocking history and *Cyclops* diel vertical migrations in alpine lakes above the timberline

Lake	Elevation (m above sea level)	Area (hectares)	Maximum depth (m)	Secchi disk transparency (m)	Fish-stocking history		<i>Cyclops</i> vertical migration in 1985		
					First stocking	Establishment of reproducing population	Range (m)	Difference in mean depth of population significant (<i>t</i> -test) at:	Variance in depth distribution significant (<i>F</i> -test) at:
CM	1,580	21	76	18	—	—	(2)	NS	NS
CG	1,620	18	51	15	1890*	—	(4)	NS	$P = 0.1$
W	1,665	34	79	17	1980	—	6	$P = 0.01$	$P = 0.002$
CP	1,722	13	50	13	1960	1970	11	$P = 0.001$	$P = 0.002$
P	1,668	8	35	7	1960	1970	7	$P = 0.001$	$P = 0.002$
Z	1,672	4	15	7	1948	1965	6	$P = 0.001$	$P = 0.02$
G	2,413	2	10	6	1700	1700	9†	$P = 0.001$	$P = 0.002$
M	1,393	35	51	11	Natural fish stock		18	$P = 0.001$	$P = 0.002$

The zooplankton community in all stocked lakes is composed of *Cyclops* and rotifers exclusively. Abbreviations of lakes: CM, Czarny nad Morskim; CG, Czarny Gasienicowy; W, Wielki; CP, Czarny w Pieciu; P, Przedni; Z, Zielony; G, Gossenköllesee; M, Morskie Oko. The last two columns represent a comparison between noon and midnight values. NS, not significant.

* Unsuccessful.

† Migration in 1962.

greatest in Lake Czarny w Pieciu, where efficient predation may extend to deeper strata because the water is clearer than in the other lakes.

An increase in the range of *Cyclops* migration since the time of stocking was also observed in the shallow Lake Zielony (Fig. 3). In 1962, a few years after the lake was first stocked and before fish had become established as a reproducing population (Table 1), migration was not detected (the difference between the noon and midnight mean depth was not significant at $P = 0.2$; F -test). However, migratory behaviour was evident in 1985. The range of migration is still much lower in Lake Zielony than in another shallow lake at Kühtai, in the Alps, which was stocked several hundred years ago (Fig. 3). The migratory range in Lake Zielony seems sufficient to evade predation as only a few copepods were found in fish guts during the summer¹⁵.

Vertical migration of *Cyclops* is probably not induced as a simple phenotypic response; the slow increase in the range of migration of the population over several generations (one generation per year) indicates that natural selection is involved. The deep Lake Morskie Oko and the shallow Lake Gossenköllesee demonstrate that hundreds or thousands of generations are necessary for the development of migratory behaviour in the population. A shorter time is probably needed for the development of migratory behaviour in a shallower lake where a lower range of migration is sufficient to allow the animals to find a safe refuge in the sediments. Factors other than predation, such

as phototoxicity^{16,17}, may be responsible for minor daily migrations, but predation is the probable cause of strong selection for a major (and metabolically costly) vertical migration in copepods.

The evolution of migratory behaviour, combined with the evolution of resistance of eggs to digestion¹⁴, might explain why *Cyclops* is the sole crustacean survivor in alpine lakes stocked with planktivorous fishes.

I thank Miss Malgosia Witeska for help in the field.

Received 14 November 1985; accepted 25 February 1986.

1. McLaren, I. A. *J. Fish. Res. Bd Can.* **20**, 685–727 (1963).
2. McLaren, I. A. *Am. Nat.* **108**, 91–102 (1974).
3. Kerfoot, W. C. *Am. Nat.* **104**, 529–546 (1970).
4. Enright, J. T. *Limnol. Oceanogr.* **22**, 856–872 (1977).
5. Zaret, T. M. & Suffern, J. S. *Limnol. Oceanogr.* **21**, 804–813 (1976).
6. Wright, D., O'Brien, W. J. & Vinyard, G. L. in *Evolution and Ecology of Zooplankton Communities* (ed. Kerfoot, W. C.) (New England University Press, 1980).
7. Stich, H.-B. & Lampert, W. *Nature* **293**, 396–398 (1981).
8. Stich, H.-B. & Lampert, W. *Oecologia* **61**, 192–196 (1984).
9. Pijanowska, J. & Dawidowicz, P. *Hydrobiologia* (in the press).
10. Gliwicz, Z. M. *Ecology* (in the press).
11. Gliwicz, Z. M. *Arch. Hydrobiol., Beih. Ergebn. Limnol.* **20**, 419–430 (1985).
12. Gliwicz, Z. M. in *Paleolimnology of Lake Biwa and the Japanese Pleistocene* (ed. Horie, S.) (Kyoto University Press, 1980).
13. Gliwicz, Z. M. *Wiad. ekol.* (in the press).
14. Gliwicz, Z. M. & Rowan, M. *Limnol. Oceanogr.* **29**, 1290–1299 (1984).
15. Dawidowicz, P. & Gliwicz, Z. M. *Envir. Biol. Fishes* **8**, 55–60 (1983).
16. Hairston, N. G. Jr *Proc. natn. Acad. Sci. U.S.A.* **73**, 971–974 (1976).
17. Luecke, Ch. & O'Brien, W. J. *Limnol. Oceanogr.* **26**, 454–460 (1981).
18. Gliwicz, Z. M. *Polskie Archiwum Hydrobiol.* **14**, 53–72 (1967).
19. Eppacher, T. *Ber. naturw.-med. Ver. Innsbruck* **56**, 31–123 (1968).

Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement

M. A. Goodale, D. Pelisson & C. Prablanc

Laboratoire de Neuropsychologie Expérimentale, INSERM, Unité 94, Bron 69500, France, and Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2

When we reach towards an object that suddenly appears in our peripheral visual field, not only does our arm extend towards the object, but our eyes, head and body also move in such a way that the image of the object falls on the fovea. Popular models of how reaching movements are programmed^{1,2} have argued that while the first part of the limb movement is ballistic, subsequent corrections to the trajectory are made on the basis of dynamic feedback about the relative positions of the hand and the target provided by central vision. These models have assumed that the adjustments are dependent on seeing the hand moving with respect to the target. Here we present evidence that a change in the position of a visual target during a reaching movement can modify the trajectory even when vision of the hand is prevented. Moreover, these dynamic corrections to the trajectory of the moving limb occur without the subject perceiving the change in target location. These findings demonstrate that (1) visual feedback about the relative position of the hand and target is not necessary for visually driven corrections in reaching to occur, and (2) the mechanisms that maintain the apparent stability of a target in space are dissociable from those that mediate the visuomotor output directed at that target.

In our experiment, four subjects were seated in turn in a darkened room with the index finger of their right hand resting on a small visual target located on a platform directly in front of them. They were instructed to move their finger as quickly and as accurately as they could to the new location of the target as it jumped from its central position to a more peripheral one. At the beginning of a trial, the subjects viewed their hand and arm through a half-reflecting mirror. The target consisted of a small light-emitting diode (LED) positioned above the mirror

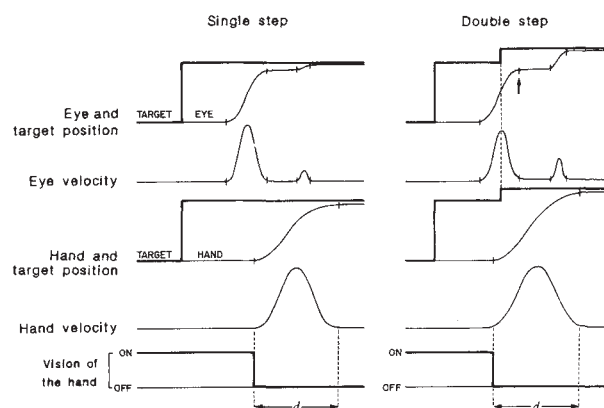


Fig. 1 Schematic diagram illustrating the spatiotemporal organization of the movements of the target, eyes and hand for trials in which single- and double-step displacements of the target occurred. The duration of the hand movement is indicated by d . The entire experiment was controlled and monitored by an on-line computer. As the diagram indicates, the second displacement of the target in double-step trials was triggered when the first saccade reached maximum velocity and vision of the hand was prevented as soon as the hand had begun to move. Thirty single- and 30 double-step trials were presented to each of the 4 subjects in random order. Randomization of the single- and double-step trials was critical to the experimental procedure. Changing the retinal feedback from a saccade in a systematic way results in a progressive and adaptive change in the gain of the saccades⁷. By presenting the subject with single- or double-step targets in random order, we were able to keep the gain of the oculomotor response constant. Moreover, even if the subjects were able somehow to detect the second jump, they would be unable to predict when such a trial would occur.

so that its virtual image appeared below the mirror on the same surface as the subject's hand. The head was fixed with a bite bar and eye movements were monitored by electro-oculography. The position of the finger at the beginning and end of a movement was derived from resistive paper that covered the surface of the platform³. The trajectory was reconstructed from a photograph of a flashing infrared LED (2 ms at 100 Hz) attached to

the finger. As soon as the subject's finger left the surface to point to the peripheral target, the subject's view of his hand and arm disappeared. It reappeared only after the target moved back to the central position, and when the hand began to move back in preparation for a new trial.

On half the trials (single step), the target jumped from its central position to a randomly determined position 30, 40 or 50 cm to the right and stayed there until the finger re-contacted the surface. On the remainder of the trials (double step), the target jumped to one of these positions, followed by a second jump to a new position 10% farther away. The second jump occurred just as the first saccade reached peak velocity. (We were taking advantage of the fact that during a reaching movement the eyes begin to move much sooner than the hand and that subjects typically make at least two saccades in foveating the target. Thus, at the end of the first saccade, they would have an opportunity to make a larger correction saccade and to modify the trajectory of their limb movement to accommodate the change in target position.) Subjects were not told that the target would sometimes change position during the first saccade. The two conditions are illustrated schematically in Fig. 1.

The results of the experiment were clear and unambiguous. As Fig. 2 illustrates, although the subjects consistently undershot the target, the distributions of their final positions in the single- and double-step conditions were separated by a distance equivalent to the size of the second displacement; in other words, the subjects were correcting their trajectory to accommodate this displacement. Moreover, the duration of a pointing movement made towards a displaced target corresponded to the duration of a movement that would have been made to a non-displaced target located at the same final position, indicating that no additional processing time was required on double-step trials (Fig. 3). The velocity profiles of the limb movement in these trials also showed no 'breakpoint' to indicate a reprogramming of the manual response.

These results strongly suggest that when rapid eye and hand movements are directed at a visual target, an initial set of signals is sent to the muscles controlling both the eye and the hand based on information about the position of the target available from the peripheral retina; after the first saccade, however, the combined input from retinal and extra-retinal signals updates the information about target position, which is then immediately used to 'fine-tune' the trajectory of the hand. If this hypothesis is correct, it becomes much clearer why the duration of a pointing movement is the same for a given position of the target, independent of whether that position was achieved by a single or a double displacement of the target. In either case, the same post-saccadic information would be used to update the motor signals controlling the trajectory of the hand (provided the second jump on a double-step trial is not too large and provided it occurs during the first saccade). The apparent correction that occurred on double-step trials was nothing more than the normal updating of the motor programming that occurs at the end of the first saccade on an ordinary trial.

Because the subjects could not see their moving limb, any new information about the target location obtained at the end of the first saccade must have been combined with information about the current position of the hand that was derived from non-visual sources such as proprioceptive feedback and/or efference-copy of the signals sent to the upper limb musculature. Whatever the mechanisms might be, it is clear that efficient modulation of visuomotor output can occur in the absence of visual feedback about limb position.

To our surprise, none of the subjects realized that the target sometimes jumped to a new location while they were reaching towards it, nor did the subjects report anything different about their reaches on double-step trials as opposed to single-step trials. Thus, they failed to perceive the change in target position even though their visuomotor output was modified to accommodate that change. Furthermore, in two additional forced-choice

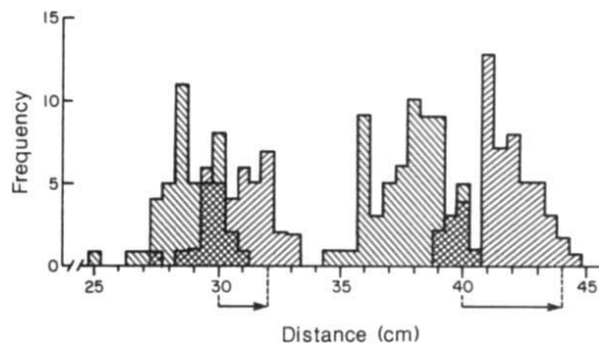


Fig. 2 Frequency distributions of the final positions of the index finger for the four subjects in single- and double-step trials for the 30- and 32-cm and the 40- and 44-cm targets. □, The frequency distributions of the final position in single-step trials. ▨, The distributions for double-step trials. The overlap of the two sets of distributions is indicated by cross-hatching. The sizes of the second displacements are indicated by the small arrows beneath the abscissa. The same pattern of undershoots was seen for the 50- and 54-cm targets. An analysis of variance yielded no significant differences between the undershoots for any of the three pairs of amplitudes.

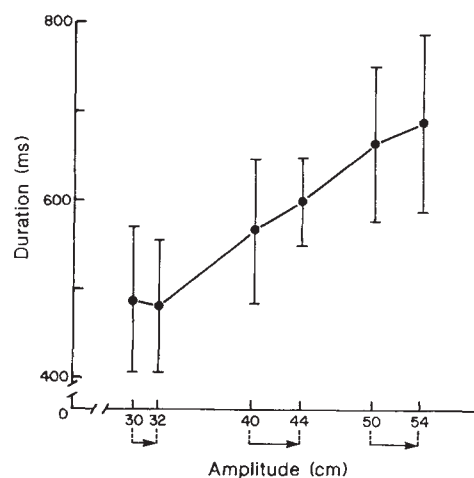


Fig. 3 Mean duration of the hand movements plotted as a function of their amplitude. Vertical lines indicate standard deviations.

experiments, subjects were unable to discriminate between double-step and single-step trials or between trials in which the target jumped away from the central fixation point during its second displacement and trials in which it jumped in the opposite direction. Their hit rate remained at chance despite the fact that they were encouraged to use any visual cue that was available. A similar dissociation between motor output and perceptual report has been observed in experiments in which subjects were asked to point towards or report verbally about a target that might or might not have changed its position⁴.

The failure to perceive the second displacement in our experiments may reflect the broad tuning of perceptual constancy mechanisms that preserve the identity of a target as its position is shifted on the retina during an eye movement. We believe that these experiments constitute the first clear demonstration (in the intact human) that the neural mechanisms mediating the perception of target position can be dissociated from those mediating visually guided reaching movements directed at that target. This dissociation is reminiscent of the behaviour of cortically blind patients who fail to perceive visual targets even though they can point towards them^{5,6}.

This research was supported in part by a grant from the MRC of Canada to M.A.G.

Received 30 December 1985; accepted 13 February 1986.

1. Keele, S. W. in *Handbook of Physiology. The Nervous System*. Vol. 2 (ed. Brooks, V.) 1391-1414 (American Physiological Society, Baltimore, 1981).
2. Paillard, J. in *Analysis of Visual Behavior* (eds Ingle, D., Goodale, M. A. & Mansfield, R. M.) 367-385 (MIT Press, 1982).
3. Bauer, J. A. Jr, Woods, G. D. & Held, R. A device for rapid recording of positioning responses in two dimensions, *Behavioral Research Methods and Instrumentation* 1, 157-159 (1969).
4. Bridgeman, B., Lewis, S., Heit, G. & Nagle, M. *J. exp. Psychol., hum. Percept. Perform.* 5, 692-700 (1979).
5. Perenin, M. T. & Jeannerod, M. *Neuropsychology* 16, 1-13 (1978).
6. Weiskrantz, L., Warrington, E., Sanders, M. D. & Marshall, J. *Brain* 97, 709-728 (1974).
7. Henson, D. B. *Vision Res.* 18, 63-67 (1978).

Reduction in number of immunostained GABAergic neurones in deprived-eye dominance columns of monkey area 17

S. H. C. Hendry & E. G. Jones

Department of Anatomy, University of California, Irvine, California 92717, USA

The primary visual cortex (area 17) of the Old World monkey is divided into alternating right- and left-eye dominance columns¹⁻³ that are highly modifiable by visual experience during a critical period in development but display little morphological or physiological plasticity during adult life^{4,5}. However, changes in immunocytochemical staining for a calcium/calmodulin-dependent protein kinase occur in visual cortical neurones of adult monkeys after brief monocular deprivation⁶ and concentrations of putative neurotransmitters or their related enzymes can be altered

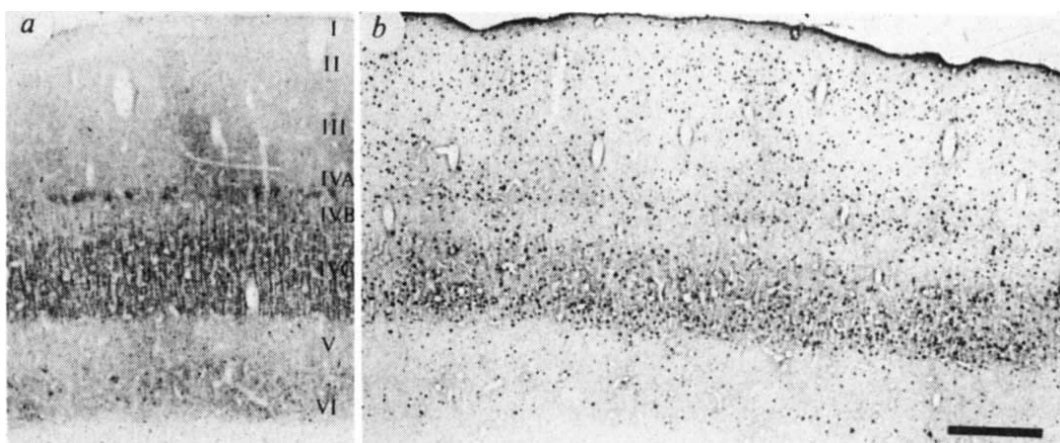
with changes in neuronal activity in other systems⁷⁻⁹. We therefore examined the effects of monocular deprivation on the immunocytochemical staining for γ -aminobutyric acid (GABA) and its synthetic enzyme, glutamic acid decarboxylase (GAD), in adult monkey area 17. The staining for GABA and GAD in neuronal somata and terminals was markedly reduced within ocular dominance columns associated with a removed or a visually deprived eye, suggesting that the GABA concentration in cortical neurones may depend on their levels of activity. Thus area 17 of adult monkeys may retain a greater degree of plasticity than previously recognized and sensory experience can profoundly affect transmitter levels, in the cortex, apparently by regulating levels of a synthetic enzyme.

In normal monkeys staining reveals GABA-positive somata in all layers of area 17, but concentrated in layers II, IVA and IVC (Fig. 1a,b). This distribution pattern agrees with that reported for neurones in monkey area 17 stained immunocytochemically for GAD¹⁰. The GABA-positive neurones possess the morphological features of non-pyramidal cells in having rounded somata and stained processes that emerge from many points along their surfaces. Preliminary electron microscopic examinations indicate that all the GABA-positive cells are neurones. The somata, particularly in layer IVC, are scattered among dense aggregates of stained punctate profiles, which are interpreted as axon terminals.

The distribution of GABA-positive somata within layers IVA and IVC appears homogeneous (Fig. 1a). In 30 consecutive 100- μ m-wide blocks through layer IVC of two normal monkeys, the mean number of GABA-positive neurones was 28.4 (minimum 26, maximum 32).

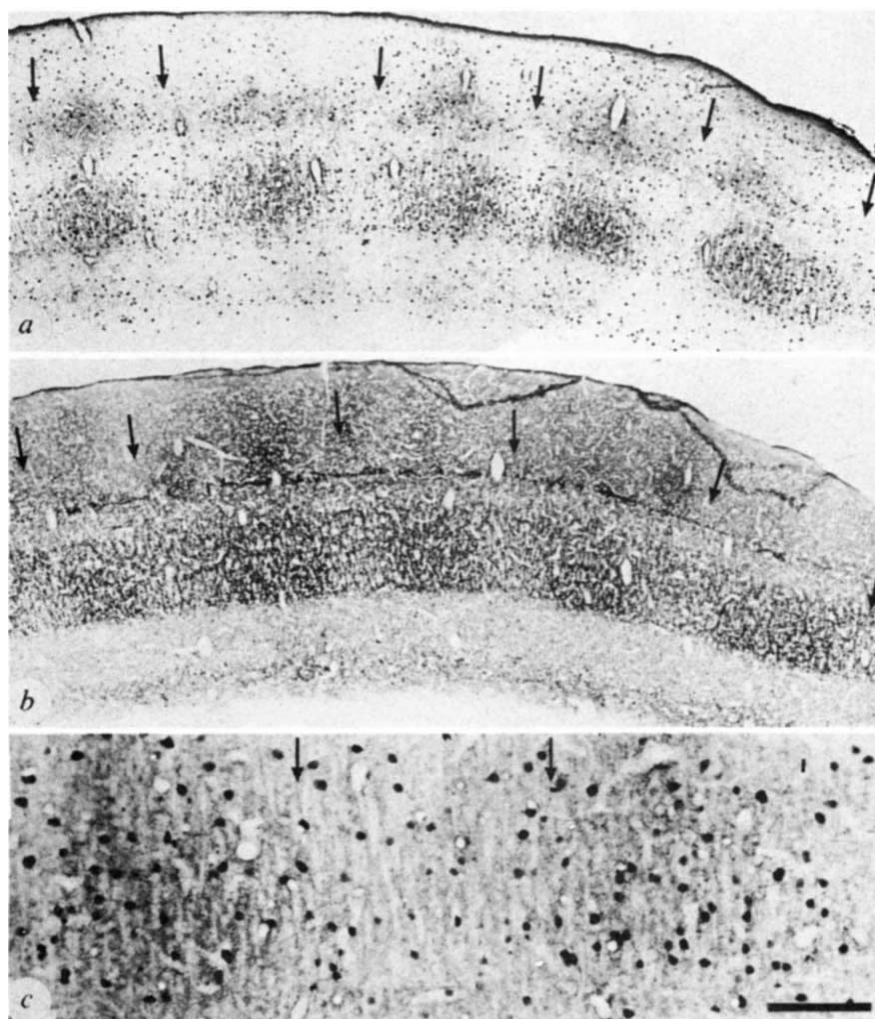
GABA staining experiments in enucleated monkeys showed that 2 weeks after removal of one eye, the staining of GABA-positive somata and terminals becomes patchy in layer IVC and is also reduced in layer IVA (Fig. 2a). Qualitatively, the supra- and infra-granular layers appear unaffected. In both layers IVA

Fig. 1 Photomicrographs of adjacent sections through area 17 of a normal monkey stained for CO (a) and GABA (b). Staining for both CO and GABA is most intense in layers IVA and IVC, with the largest number of GABA-positive cells found in the deep half of layer IVC. GABA-positive cells are also scattered throughout the remaining layers and are relatively dense in layer II and the superficial half of layer III. Within each layer the distribution of GABA-positive cells is homogeneous. Scale bar, 400 μ m.



Methods. Eight young adult cynomolgus monkeys (*Macaca fascicularis*) were used in this study. Two were normal; four had one eye removed under barbiturate anaesthesia 2, 3 or 6 weeks before they were killed, and two had the lids of one eye sutured together 7 or 11 weeks before they were killed. The ages of most of the monkeys were unknown, but their weights (3-4.5 kg) indicate they were at least 2 years old. The monkey monocularly deprived for 11 weeks was 22 months old when killed. All animals were perfused with 2% paraformaldehyde and 0.2% or 0.4% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4). Alternate sections through area 17 were cut frozen at 15 or 30 μ m. Most of the thinner sections were incubated in anti-GABA antiserum (Immunonuclear, Inc.) diluted 1:1,500 or 1:3,000 in 0.1 M phosphate buffer containing 3% normal goat serum and 0.2% Triton X-100, or in anti-GAD antiserum (supplied by Dr D. Schmechel) at 1:1,500 in the same buffer containing 3% normal rabbit serum for 12-18 h at 4°C. They were then processed by the peroxidase/anti-peroxidase method¹⁷ or the avidin-biotin method¹⁸. Staining was qualitatively and quantitatively identical with the two fixation and immunocytochemical procedures used. The thicker sections were processed histochemically for cytochrome oxidase (CO)¹⁹. Some thinner sections were stained with thionine. Control sections, incubated in primary antiserum adsorbed with an excess (5-15 μ g ml⁻¹) of GABA conjugated to bovine serum albumin or free GABA (15-30 μ g ml⁻¹) contained very light, diffuse staining only. The numbers of GABA-positive somata in 100- μ m-wide blocks through the full dorso-ventral extent of layer IVC were counted in normal monkeys, in deprived- or removed-eye columns and in non-deprived- or intact-eye columns. Counts were made from 30 blocks in each of three sections from the two normal monkeys, two enucleated monkeys (2 and 3 weeks survival post-enucleation) and the two lid-sutured monkeys. The same procedure was used in these animals to count the total number of thionine-stained cells (with nuclei in the section) in radial traverses through the full thickness of area 17 and through layer IVC alone. For counts in the manipulated monkeys, the adjacent CO-stained section was used to locate the two sets of columns in area 17.

Fig. 2 Photomicrographs of sections through area 17 of a monkey that had one eye surgically removed 2 weeks before killing. The GABA staining (*a*) in layers IVA and IVC is broken up into large (500–700 μm wide) intensely stained patches and smaller (250–400 μm wide) lightly stained patches. Comparison with the CO staining in a neighbouring section (*b*) reveals that the intense GABA staining corresponds to intact-eye dominance columns and the light staining to removed-eye dominance columns (arrows). A photomicrograph of layer IVC at higher magnification (*c*) shows the reduction in GABA-stained somata and terminals in a removed-eye dominance column (between arrows). Scale bar, 500 μm in *a*, *b* and 150 μm in *c*.



density as in normal monkeys, alternate with patches containing fewer and more lightly stained somata and terminals (Fig. 2*a*, *c*). Most of the intensely stained patches are larger (500–700 μm wide) than the lightly stained patches (250–400 μm wide). No differences in the sizes of the dark and light GABA patches were seen with increasing survival times after enucleation. Counts of immunoreactive somata in layer IVC of two enucleated monkeys (2 and 3 weeks post-enucleation) show that the number of cells in the densely stained patches is 27.7 ± 2.9 (mean $\pm$ s.d.) while that in the lightly stained patches is 15.2 ± 2.7 . The difference is statistically significant ($t = -18.67$; $P < 0.0005$). There was no significant difference between the two monkeys. In these enucleated monkeys, the total number of thionine-stained neurones in the removed-eye columns did not differ significantly from the number in the intact-eye columns, nor did it differ from the number in normal monkeys.

Cytochrome oxidase (CO) staining in layers IVA and IVC is also inhomogeneous in enucleated monkeys, with alternating intensely and lightly stained patches, each 400–600 μm wide (Fig. 2*b*), representing the ocular dominance columns associated with the intact and the removed eye, respectively^{10,11}. Comparison of Fig. 2*a* and *b* demonstrates that the patches intensely stained for GABA coincide with patches intensely stained for CO, but the dark GABA patches frequently extend beyond the borders of the dark CO patches. The lightly stained GABA and CO patches also coincide, but light GABA patches generally occupy only the central half to two-thirds of the light CO patches. No effect on GABA staining was apparent in neurones above or below layer IV.

Figure 3 shows sections through area 17 of a monocularly deprived monkey. The lids of one eye were sutured together 11 weeks before the animal was killed and sections were cut parallel to the opercular surface of the occipital lobe. In the section stained for CO (Fig. 3*b*), the alternating darkly and lightly stained ocular dominance columns in layer IVC appear as elongated bands, as does the GABA staining in the adjacent section (Fig. 3*a*). By superimposing the same blood vessels in the two sections, the intensely stained GABA bands are found to coincide with the intensely stained CO bands. The two sets of lightly stained bands also coincide, but the decreased GABA staining usually occupies only the central half to two-thirds of a deprived-eye band. The numbers of stained somata and terminals are significantly reduced in the deprived-eye bands (Fig. 3*c*; mean of non-deprived columns, 28.4 ± 3.4 ; deprived columns, 18.1 ± 3.1 ; $t = 17.31$, $P < 0.0005$). Total neuronal density in the deprived columns is equal to that in the non-deprived columns and in the normal monkeys. The eyes of both deprived animals were histologically normal and the cells in the deprived laminae of the dorsal lateral geniculate nuclei displayed no shrinkage.

Immunocytochemical staining for GAD was also changed in the monkey deprived for 11 weeks. As with GABA, the GAD staining in layers IVA and IVC consists of alternating dark and light bands that correspond to non-deprived and deprived ocular dominance columns, respectively (Fig. 3*d*).

These results show that eye removal or monocular deprivation reduces the number of immunocytochemically stained GABA and GAD somata and terminals in deprived-eye dominance

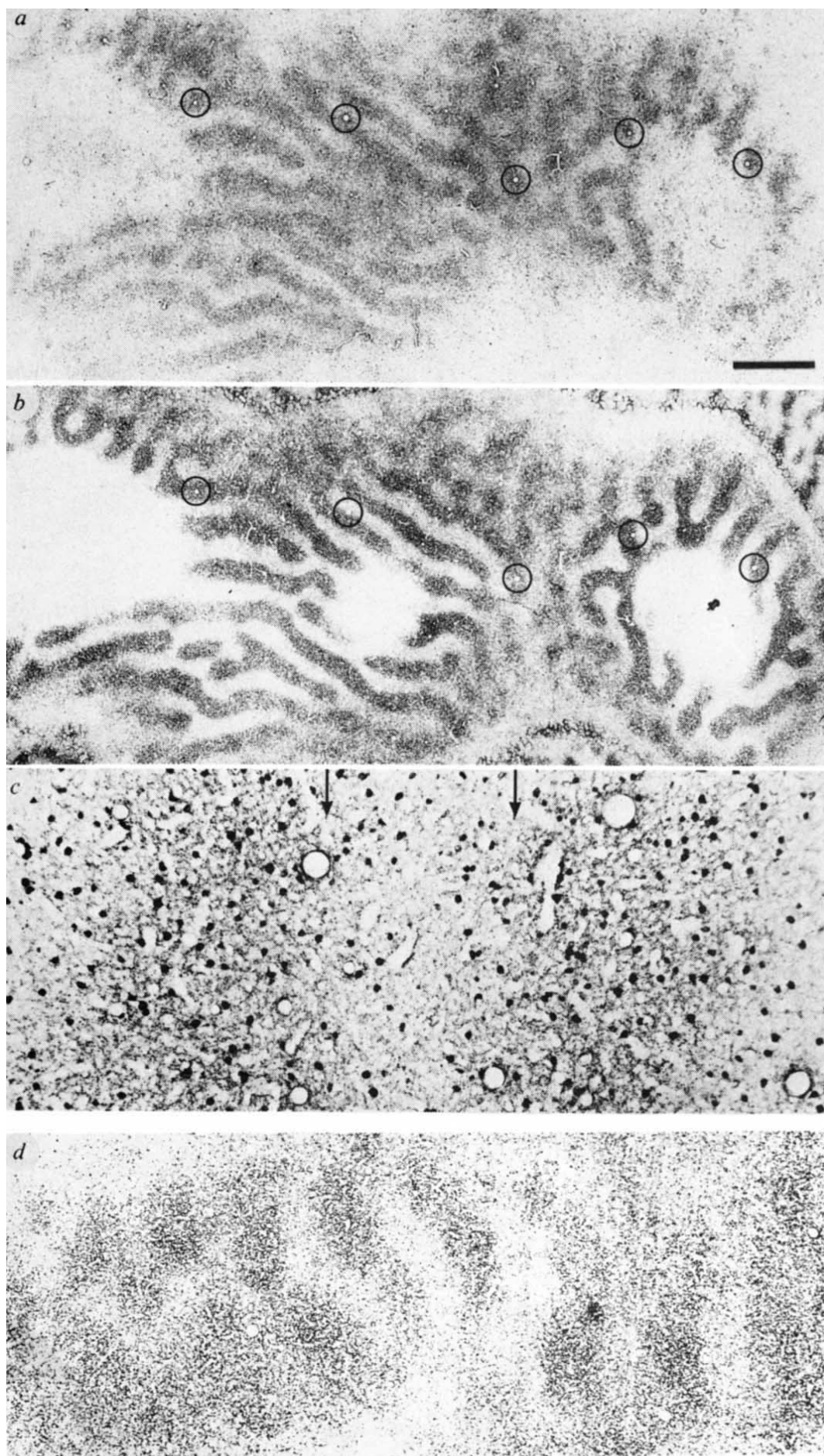


Fig. 3 Photomicrographs of tangential sections through the operculum of the occipital lobe taken from a monkey that had the lids of one eye sutured together 11 weeks before it was killed. The GABA (*a*) and CO (*b*) staining in layer IVC of area 17 is distributed in elongated bands. By superimposing the same radially oriented blood vessels (circles) in these adjacent sections, the two sets of intensely stained bands and the two sets of lightly stained bands are found to coincide, demonstrating that the GABA staining declines in the deprived-eye dominance columns. At higher magnification (*c*) the density of GABA-positive somata and terminals in a deprived eye-column (between arrows) is found to be much lower than that in the two flanking intact-eye columns. GAD staining in layer IVC is also distributed in alternating dark and light bands (*d*) that coincide with dark and light CO bands in the adjacent section (not shown). Scale bar, 2 mm in *a*, *b*; 100 μ m in *c*; 750 μ m in *d*.

columns in monkey area 17. Monocular deprivation in kittens reportedly produces no changes in the immunostaining for GAD in area 17 neurones¹². The differences between these results may be related to differences in the species or methods used.

We detected no reduction in the total neuronal density in area 17 of the enucleated and lid-sutured monkeys. It appears, then, that a mechanism other than the degeneration of GABA-positive somata and axon terminals reduces the numbers of these stained elements in deprived-eye columns. We interpret this reduction to be due to a decline in the GABA and GAD concentrations within individual neurones and terminals, to the point where the substances are no longer detectable by immunocytochemistry. The reduced staining for GABA and GAD probably does not reflect a general reduction in synthetic capability by deprived neurones since the immunostaining for calcium/calmodulin-dependent protein kinase type II is increased in neurones within deprived-eye columns⁶.

The reductions in CO staining of deprived-eye columns was much greater in the present study than that reported previously following comparable periods of eyelid closure in adult monkeys¹³, and more closely resembled the deprivation produced by intravitreal injection of tetrodotoxin¹⁴. Technical modifications probably account for the difference in CO staining; the difference is apparently not due to accidental damage to the sutured eyes since the eyes and lateral geniculate nuclei were normal in the deprived animals.

Because of the influence of intracortical GABA mechanisms on receptive field structure in the visual cortex¹⁵, it seems reasonable to expect that the reduction of GABA levels in cortical intrinsic neurones and axon terminals should produce changes in the physiological activity of their target cells. These changes

could account for the functional expansion of intact-eye dominated columns that has been reported for area 17 of one enucleated adult monkey⁵, even though no expansion of thalamocortical axons related to an intact eye has been found anatomically in adults^{4,5}. Similar expansions have not been reported in adult monkeys after prolonged periods of eye closure^{4,5,16} even though our results indicate that this manipulation does reduce GABA and GAD levels. As these data appear to show a regulation of cortical transmitter levels by sensory experience, it will now be important to determine the exact mechanism of this regulation.

This work was supported by NIH grant NS 21317 to E.G.J. and a Sloan Foundation Fellowship to S.H.C.H. We thank the Doheny Eye Foundation, University of Southern California Medical School for providing three of the monkeys.

Received 26 August 1985; accepted 20 February 1986.

- Hubel, D. H. & Wiesel, T. N. *J. comp. Neurol.* **146**, 421-450 (1972).
- Wiesel, T. N., Hubel, D. H. & Lam, D. M. K. *Brain Res.* **79**, 273-279 (1974).
- Kennedy, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 4230-4234 (1976).
- Hubel, D. H., Wiesel, T. N. & LeVay, S. *Phil. Trans. R. Soc. B278*, 377-409 (1977).
- LeVay, S., Wiesel, T. N. & Hubel, D. H. *J. comp. Neurol.* **191**, 1-51 (1980).
- Hendry, S. H. C. & Kennedy, M. B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1536-1540 (1986).
- Mollinoff, P. B., Brimijoin, S., Weinshboum, R. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **66**, 453-457 (1970).
- Kessler, J. A. & Black, I. B. *Brain Res.* **234**, 182-194 (1982).
- Baker, H., Kawano, T., Margolis, F. L. & Joh, T. H. *J. Neurosci.* **3**, 69-78 (1983).
- Hendrickson, A. E., Hunt, S. P. & Wu, J.-Y. *Nature* **292**, 605-607 (1981).
- Horton, J. C. & Hubel, D. H. *Nature* **292**, 762-764 (1981).
- Bear, M. F., Schmechel, D. E. & Ebner, F. F. *J. Neurosci.* **5**, 1262-1275 (1985).
- Horton, J. C. *Phil. Trans. R. Soc. B304*, 199-253 (1984).
- Wong-Riley, M. & Carroll, E. W. *Nature* **281**, 572-574 (1984).
- Sillito, A. M. in *The Cerebral Cortex* Vol. 2 (eds Jones, E. G. & Peters, A.) 91-117 (Plenum, New York, 1984).
- Blakemore, C., Vital-Durand, F. & Garey, L. J. *Proc. R. Soc. B213*, 399-423 (1981).
- Sternberger, L. A. *Immunocytochemistry* (Wiley, New York, 1979).
- Hsu, S. M. & Raine, L. J. *Histochem. Cytochem.* **29**, 1349-1353 (1981).
- Wong-Riley, M. T. T. *Brain Res.* **171**, 11-28 (1979).

GABA and GAD immunoreactivity of photoreceptor terminals in primate retina

Yozo Nishimura, Michael L. Schwartz & Pasko Rakic*

Section of Neuroanatomy, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06519, USA

Within the vertebrate retina, two types of photoreceptor cells—the rods and cones—transduce visual signals and convey this information through synapses with bipolar and horizontal cells. Although the neurotransmitter at these first-order synapses has not been identified, electrophysiological studies^{1,2,3} suggest that it might be excitatory. In the present study, however, we have found photoreceptor terminals in the rhesus monkey retina which are immunoreactive with antibodies to either γ -aminobutyric acid (GABA) or L-glutamic acid decarboxylase (GAD, an enzyme involved in the synthesis of GABA). In the perifoveal region of the retina, approximately 25% of presynaptic profiles having ultrastructural characteristics of either rod or cone terminals are immunoreactive with one or the other antibody. This evidence for a putatively inhibitory neurotransmitter in photoreceptor terminals challenges present understanding of retinal synaptic function.

Our recent light microscopic immunocytochemical analysis of monkey retina using antibodies to either GABA or GAD consistently revealed a narrow band of reactivity in the part of the outer plexiform layer (OPL) adjacent to the photoreceptor cells⁴. As previous studies of the mammalian retina have not resolved the nature of the GABA-containing elements in this synaptic layer, we decided to determine the precise localization and source(s) of this immunoreactivity at the electron microscope level.

Retinas from 22 monkeys (*Macaca mulatta* and *Macaca fascicularis*) were fixed by perfusion with various combinations of aldehydes in phosphate-buffered saline (PBS). The combination of 4% paraformaldehyde and 0.1% glutaraldehyde consistently gave the greatest sensitivity to both GABA and GAD antibodies, with low background staining and good ultrastructural preservation. The retinas were embedded in agar, 50- μ m-thick Vibratome sections were cut through the perifoveal region and the sections were then reacted with either anti-GABA or anti-GAD antiserum using the avidin-biotin-peroxidase conjugation method⁵. The reacted retinas were postfixed in osmium and embedded in Epon-Araldite. Prior to ultrathin sectioning, two to three 1- μ m-thick sections from the surface of the block were discarded to ensure that nonspecific peroxidase precipitate, which may be found at the surface of a block, was not included in the sections to be examined under the electron microscope. This nonspecific reactivity was uniformly absent in sections deeper than 2 μ m. Thus, all ultrathin (70 nm) sections were taken at a distance of 2 to 7 μ m from the surface of the block. In some specimens, up to 30 serial ultrathin sections were cut in order to reconstruct the terminals and determine their origins. Quantitative analyses were performed on three retinas using sections situated 2-4 μ m from the surface of the block. These sections were selected because the penetration of the antibodies gradually diminishes in deeper sections. To determine the relationship between labelled terminals and adjacent elements of the OPL, one block was cut slightly oblique to the plane which is parallel to the retinal laminae.

In accord with previous studies on the vertebrate retina⁶⁻¹⁶, light microscopic analysis of the Vibratome sections stained with either the anti-GABA or anti-GAD antiserum revealed intense immunoreactivity in some amacrine cells and in the neuropil of the inner plexiform layer (IPL; Fig. 1). In addition, we confirmed our recent finding of labelled horizontal cells⁴, and noted slight immunoreactivity in some cells of the outer nuclear layer (ONL). However, most importantly for the purposes of the present study, we consistently detected a dark band of globular immuno-

* To whom correspondence should be addressed.

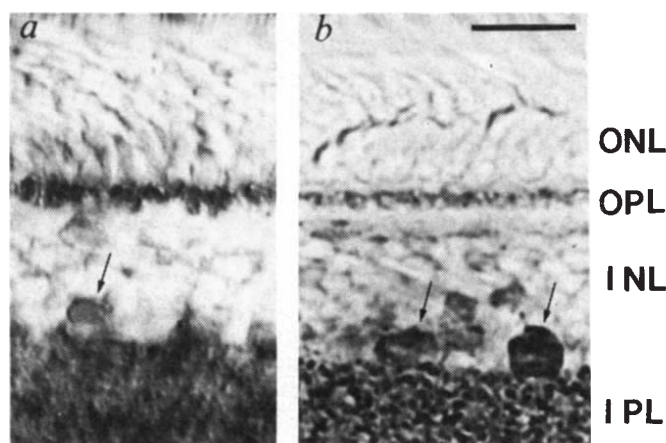


Fig. 1 *a*, Light micrograph of the perifoveal region of primate retina reacted with anti-GAD antiserum. *b*, The same region of the retina reacted with anti-GABA antiserum. Both preparations show a similar distribution of immunoprecipitate. Note that in addition to amacrine cells (arrows), immunoreactivity is also localized in the portion of the outer plexiform layer (OPL) adjacent to the photoreceptors. ONL, outer nuclear layer containing photoreceptor cells; INL, inner nuclear layer; IPL, inner plexiform layer. Scale bar, 10 μ m.

Methods. The anti-GABA antiserum was generated in rabbit against GABA conjugated to bovine serum albumin (BSA; Immuno-Nuclear Corporation) and used at a dilution of either 1:2,500 or 1:5,000 with 1% normal goat serum in PBS. This antibody does not cross-react with a variety of amino acids and peptides, including alanine, aspartate, glutamate, glycine, taurine, substance P, serotonin, tyrosine and Met-enkephalin²⁸. Furthermore, pretreatment with GABA or GABA conjugated to BSA (20 μ g ml⁻¹) results in the loss of immunoreactivity in cerebellar tissue²⁸. The anti-GAD antiserum was generated in sheep against GAD purified from rat brain synaptosomes and the specificity of this antibody has been verified biochemically and histochemically^{29,30}. This antiserum was used at a dilution of 1:2,000 with 1% normal donkey serum in PBS. We performed several additional control experiments using retinal tissue, including incubation with normal rabbit serum (for GABA) or normal sheep serum (for GAD) instead of the primary antiserum or incubation in diluted anti-GABA serum (1:5,000) that had been pre-absorbed with GABA (100 μ g ml⁻¹; Sigma). No immunoreactivity was seen in the OPL in any of these controls.

precipitate in the outer half of the OPL (Fig. 1). Electron microscopic examination of tissue reacted with either the anti-GABA or anti-GAD antiserum revealed a considerable number of immunoreactive synaptic terminals within this band. The dense immunoprecipitate was present in some terminal profiles, while other terminals in the same sections were completely devoid of either GABA or GAD labelling (Figs 2, 3).

Immunoreactivity in the OPL was predominantly localized in a class of small (1.5–2.5 μ m) synaptic terminals, some of which were unambiguously identified as rod spherules on the basis of their size and ultrastructure. One of the best criteria for their identification as photoreceptor terminals was the presence of electron-dense synaptic ribbons surrounded by synaptic vesicles (Figs 2, 3*a, b*). In some instances where ribbons were not present, immunoreactive profiles were followed in serial sections from the presynaptic terminal to the ribbon or the connecting fibre in the receptor cell layer (see, for example, Fig. 3*c*). However, only a few photoreceptor cell perikarya were immunoreactive and these were only lightly stained, providing a possible explanation for the failure of previous immunocytochemical studies to detect GABA and GAD in these cells at the light microscopic level. The paucity of immunoprecipitate in the soma of the photoreceptor may reflect the very thin perinuclear cytoplasm of these cells, or rapid transport of the transmitter to the terminals where it may be continuously released. Both the GABA- and GAD-positive cells participated in typical synaptic triads

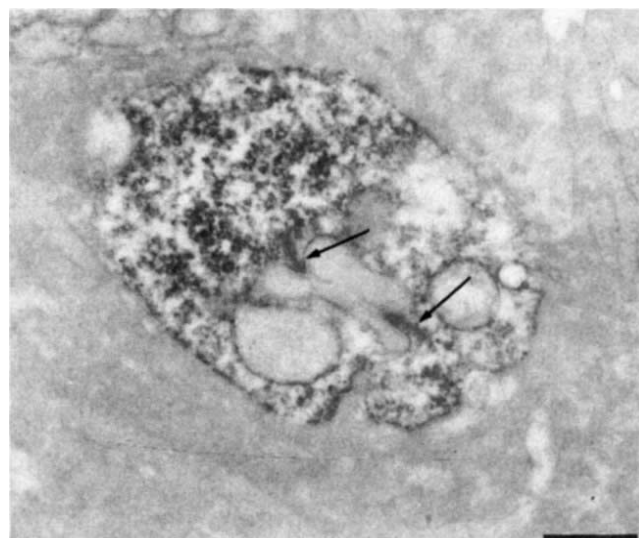


Fig. 2 Electron micrograph of a GAD-immunoreactive rod terminal in the OPL of a retina sectioned parallel to the retinal surface. The nature of the terminal can be identified on the basis of its shape, position and the presence of the characteristic synaptic ribbon (arrows). Note the absence of any immunoreactivity in the cellular elements surrounding the GAD-positive terminal. Scale bar, 1 μ m.

with processes of horizontal cells and bipolar or biplexiform cells which invaginated the photoreceptor bouton^{17,18}.

In three cases in which the labelling pattern was evaluated quantitatively, an average of 25.5% ($n=556$) of the profiles with ultrastructural characteristics of rod spherules were found to be immunoreactive. With very few exceptions, immunoreactive terminals were surrounded by unstained elements (Fig. 3). This result was verified on the specimen cut parallel to the retinal surface which displayed solitary immunopositive photoreceptor terminals surrounded by a number of reaction-free elements, which suggests that GABA-containing photoreceptor terminals may form a mosaic pattern interspersed with GABA-negative terminals. Some cone terminals, identified by their wide-based pedicles (5.0–6.5 μ m), multiple indentations, synaptic ribbons and multiple junctions, were also immunoreactive (Fig. 3*d*). Because of the lower number of cone photoreceptors in the perifoveal region of the primate retina, stained cone terminals were encountered less frequently than stained rod terminals. Nevertheless, 26.4% of cones ($n=163$) were immunoreactive in the three specimens used for counting. (Note that current methods for quantifying immunoreactive profiles at the electron microscope level provide minimal rather than maximal values for the occurrences of an antigen in a population of elements.)

No transmitter substance for photoreceptor cells in the primate retina has been fully identified and characterized, although it has been suggested that glutamate or aspartate may serve as neuroactive substances in rods and cones of several other vertebrate species^{9,14,19,20}. In the present study, we have presented evidence that at least as one-quarter of the photoreceptors in the monkey retina are immunoreactive with GAD and GABA. The presence of GAD and GABA in the same proportion of terminals indicates that GABA is not only accumulated, but also synthesized by the photoreceptors; this result was unexpected as electrophysiological data indicate that photoreceptor terminals probably exert influence on their postsynaptic target by interruption of the flow of an excitatory substance^{1–3}. It should be emphasized, however, that studies in invertebrates indicate that GABA may act in either an excitatory or inhibitory way depending on the type of receptor and state of the postsynaptic elements^{21,22}. Further, recent studies of the skate retina in tissue culture have revealed that application of GABA pro-

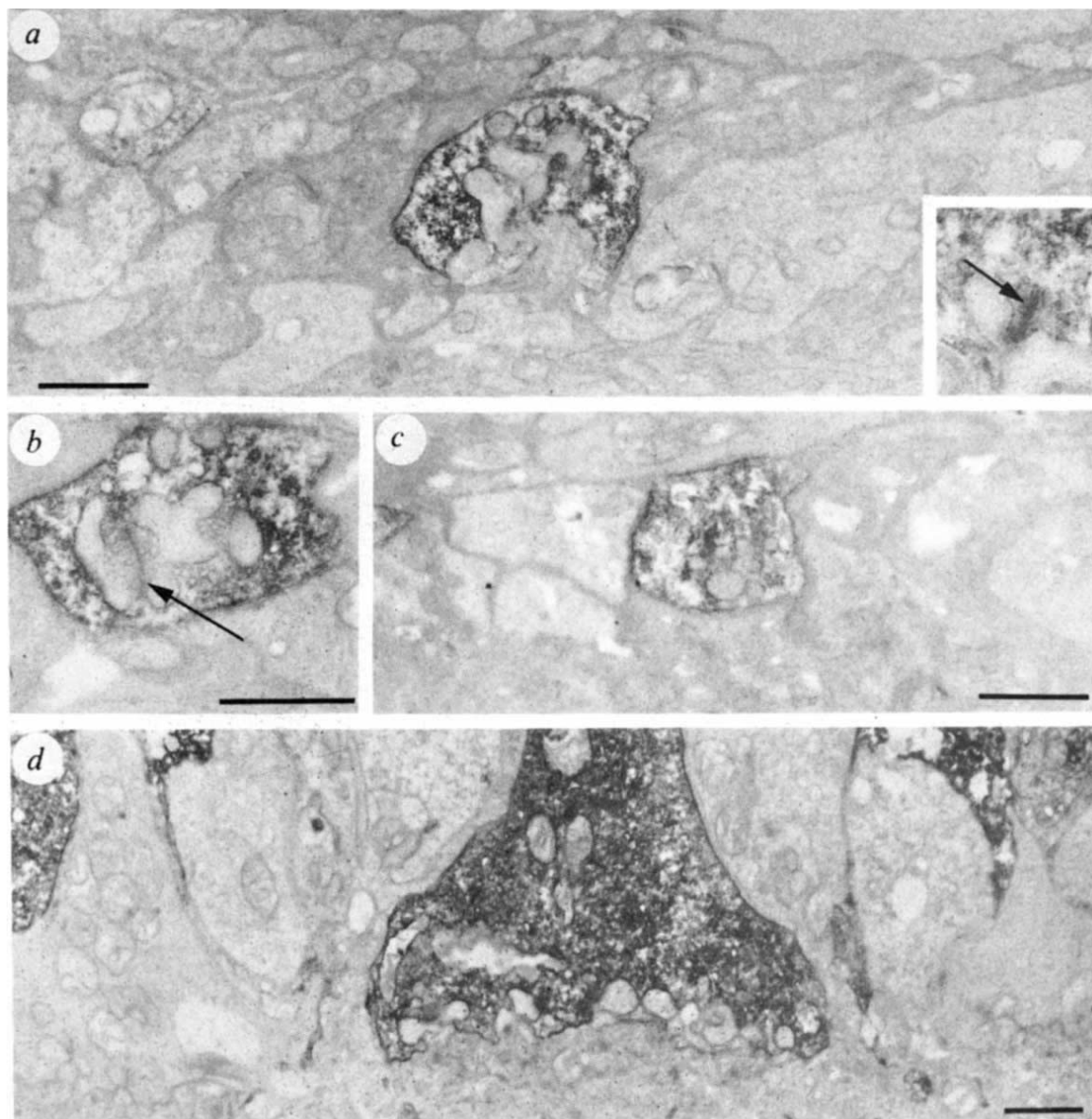


Fig. 3 Electron micrographs of immunoreactive photoreceptor terminals in the perifoveal region of the outer plexiform layer. *a*, GAD-positive synaptic terminal with the characteristics of a rod spherule. Inset, synaptic ribbon in a GAD-positive terminal. *b*, Another example of a GAD-positive rod terminal containing a typical synaptic ribbon (arrow). *c*, A GABA-positive rod terminal. *d*, A GABA-positive cone terminal with the characteristic flat cone-shaped pedicle and several invaginations forming synaptic contacts with bipolar and horizontal cells. Scale bar, 1 μ m.

duces both a depolarization of horizontal cells and a hyperpolarization of bipolar cells^{23,24}. In view of the sensitivity of horizontal and bipolar cells to GABA, and the ability of this transmitter to elicit responses of opposite polarity by acting on membrane receptors controlling different ionic channels, its presence in photoreceptor terminals may be functionally more relevant than hitherto recognized.

Although the amphibian retina contains two types of rods which differ in their morphological and physiological properties^{25,26}, mammalian rods and cones have not been previously divided into subclasses. Our findings reveal that in primates, rods and cones can be divided into at least two types: one type containing GABA or its synthesizing enzyme GAD, the other type containing neither. The GABA-positive terminals form a synaptic triad which may affect both the local circuits within the OPL via horizontal cells, as well as centrally directed output pathways via bipolar cells. One can envisage several possible ways in which GABA-containing photoreceptors could affect the processing of visual input. Perhaps the most obvious scenario

is that the two classes of photoreceptors, GABA-positive and GABA-negative, may provide a synaptic mechanism for the two functional classes of receptive fields—'on centre' and 'off centre'. It may not be coincidental that at low light levels, rods in the perifoveal region of the human retina tonically inhibit the adjacent cone pathway²⁷. At present, it is not clear whether GABA-containing photoreceptors synapse with only one type of bipolar cell or whether they modulate the activity of adjacent photoreceptor terminals via their influence on horizontal cells. However, the identification of the putative neurotransmitter GABA in at least one type of photoreceptor terminal opens up a promising new avenue of investigation for expanding our understanding of the mechanism of visual information processing in the first-order synapses of the primate retina.

This work was supported by the US PHS. Antiserum to GAD was supplied through the Laboratory of Clinical Science, National Institute of Mental Health, where it was developed by Drs Wolfgang H. Oertel, Donald E. Schmechel and Marcel Tappaz under the supervision of Dr Irwin J. Kopin.

Received 25 November 1985; accepted 3 March 1986.

1. Trifunov, Y. A. *Biophysika* **13**, 809-817 (1968).
2. Murakami, M., Ohtsu, K. & Ohtsuka, T. *J. Physiol., Lond.* **227**, 899-913 (1972).
3. Kaneko, A. & Shimazaki, H. *Cold Spring Harb. Symp. quant. Biol.* **40**, 537-546 (1976).
4. Nishimura, Y., Schwartz, M. L. & Rakic, P. *Brain Res.* **359**, 351-355 (1985).
5. Hsu, S. M., Raine, L. & Fanger, H. *J. Histochem. Cytochem.* **29**, 557-580 (1981).
6. Brecha, N. C., Eldred, W., Kuljis, R. O. & Karten, H. *J. Prog. retinal Res.* **3**, 185-226 (1984).
7. Blanks, J. C. & Roffler-Tarlov, S. *Expl Eye Res.* **35**, 573-584 (1982).
8. Brandon, C., Lam, D. M. K. & Wu, J. Y. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3557-3561 (1979).
9. Brecha, N. C. in *Chemical Neuroanatomy* (ed. Emson, P. C.) 85-129 (Raven, New York, 1983).
10. Ehinger, B. *Expl Eye Res.* **25**, 221-234 (1977).
11. Freed, M. A., Nakamura, Y. & Sterling, P. *J. comp. Neurol.* **219**, 295-304 (1983).
12. Pourcho, R. G. *Brain Res.* **198**, 333-346 (1980).
13. Vaughn, J. E. *et al. J. comp. Neurol.* **197**, 113-127 (1981).
14. Brunn, A. & Ehinger, B. *Expl Eye Res.* **19**, 435-447 (1974).
15. Lam, D. M. K. *et al. Nature* **278**, 565-567 (1979).
16. Marc, R. E., Stell, W. K., Bok, D. & Lam, D. M. K. *J. comp. Neurol.* **182**, 221-246 (1978).
17. Kolb, H. *Phil. Trans. R. Soc.* **258**, 261-283 (1970).
18. Mariani, A. P. *Int. Rev. Cytol.* **86**, 285-319 (1984).
19. Ariel, M. *et al. Brain Res.* **295**, 179-183 (1984).
20. Mosinger, J. L. & Altschuler, R. A. *J. comp. Neurol.* **233**, 255-268 (1985).
21. Blackenship, J. E., Watchel, H. & Kandel, E. R. *Neurophysiology* **34**, 76-92 (1971).
22. Yarowsky, P. J. & Carpenter, D. O. *Brain Res.* **144**, 75-94 (1978).
23. Dowling, J. E. *et al. Vision Res.* **23**, 421-432 (1983).
24. Lasater, E. M., Dowling, J. E. & Ripps, H. *J. Neurosci.* **4**, 1966-1975 (1984).
25. Kuhne, W. in *On the Photochemistry of the Retina and on Visual Purple* (ed. Foster, M.) 1-94 (Macmillan, London, 1878).
26. Nilsson, S. E. G. *J. Ultrastruct. Res.* **10**, 390-416 (1964).
27. Goldberg, S. H., Frumkes, T. E. & Nygard, R. W. *Science* **221**, 180-181 (1983).
28. Maley, B. & Newton, B. W. *Brain Res.* **330**, 364-368 (1985).
29. Oertel, W. H., Schmechel, D. E., Tappaz, M. L. & Kopin, I. J. *Neuroscience* **6**, 2689-2700 (1981).
30. Oertel, W. H. *et al. Neuroscience* **6**, 2701-2714 (1981).

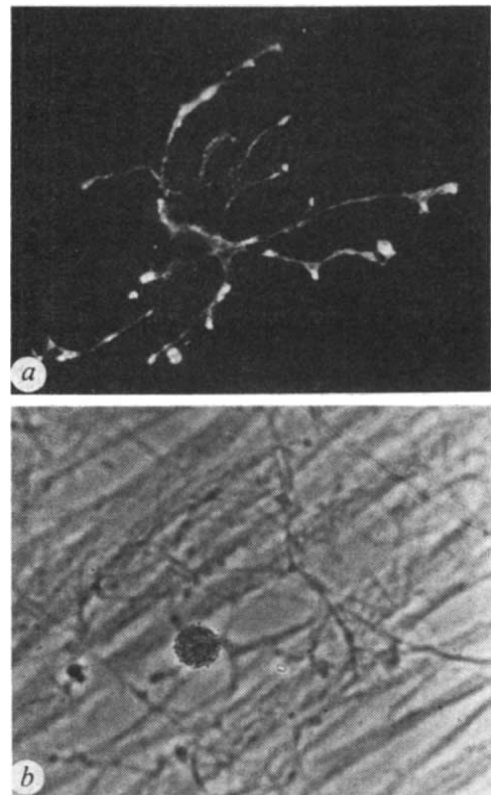


Fig. 1 Adult galC⁺ ODC which has incorporated thymidine in a neurone-containing culture. *a*, Fluorescence image of a cell given 0.1 $\mu\text{Ci ml}^{-1}$ ³H-thymidine 24 h before surface immunostaining with monoclonal anti-galC (ascites, 1:50 dilution) and rhodamine-conjugated goat anti-mouse IgG (Cappel, 1:100 dilution), fixation with 5% acetic acid/95% ethanol and autoradiography with Kodak NTB2 emulsion. *b*, Phase-contrast image of the cell in *a* focused to reveal grains over the nucleus. Note the relationship of the cell and its processes to the underlying neurites, and its relative isolation from other non-neuronal cells. $\times 415$.

Methods. The rat DRG cultures were prepared as described elsewhere⁷. CNS cells (including ODCs) were obtained from the lumbosacral spinal cord of 4-month-old female rats (Sprague-Dawley). The cord tissue was stripped of meningeal membranes, rinsed and incubated in 0.25% trypsin and 20 $\mu\text{g ml}^{-1}$ DNase in Earle's balanced salt solution for 45 min at 35 °C, and dissociated by trituration through a small-bore pipette. After filtration, the cells were separated from myelin and cell debris on a Percoll gradient, and added in suspension to the neuronal cultures or to empty collagen-coated dishes at $1-5 \times 10^3$ cells per culture. ³H-thymidine was administered 8 days after the addition of the CNS cells. The culture medium consisted of 10% human placental serum in Eagle's minimum essential medium with $\sim 50 \mu\text{g ml}^{-1}$ crude nerve growth factor.

mmol⁻¹) and monitoring incorporation by autoradiography (Fig. 1), or by fixing the cultures and scanning for the presence of mitotic figures (Fig. 2). In one experiment the cells were labelled with anti-galC antibody before exposure for 2 h to 1.0 $\mu\text{Ci ml}^{-1}$ of ³H-thymidine, and the cultures were then immediately fixed and processed for autoradiography (Fig. 3). As illustrated in Figs 1 and 2, both ³H-thymidine incorporation and mitotic figures were observed in galC⁺ ODCs in cultures containing neurones. Thus, we are confident that the thymidine labelling observed reflects cell division rather than DNA repair following the trauma of cell isolation. Figure 3 illustrates an ODC labelled with anti-galC before the addition of ³H-thymidine. Thymidine incorporation by ODCs after prelabelling with galC antibody rules out the possibility that the dividing ODCs acquired galC after completing DNA synthesis. As shown

Evidence that axons are mitogenic for oligodendrocytes isolated from adult animals

Patrick M. Wood & Richard P. Bunge

Department of Anatomy and Neurobiology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA

Recovery from demyelinating events in the adult central nervous system (CNS) may depend in some situations on the generation, by cell division, of new oligodendrocytes (ODCs)¹. Adult ODCs, identified by morphological criteria, have been shown to divide *in vivo* in response to induced demyelination or to physical trauma^{2,3}, although the factors that elicit this response are unknown. Adult ODCs do not proliferate *in vitro* under normal conditions (ref. 4 and P.M.W., unpublished observations). Encouraged by observations that axons provide a mitogenic stimulus for Schwann cells^{5,6} and support the proliferation of embryonic ODC progenitors⁷, we have studied the response of adult ODCs, identified by immunocytochemical criteria, to axons of dorsal root ganglion neurones in culture. We have now found that adult ODCs proliferate *in vitro* when neurones are present but not when neurones are absent, suggesting that neurones can elicit ODC division under the conditions prevailing in our culture system.

Experiments were constructed by adding suspensions of dissociated adult spinal cord cells to pure cultures of dissociated dorsal root ganglion neurones prepared as described previously⁷. At the time of addition of the CNS glial cells, these cultures consisted of a large, centrally located, loosely arranged cluster of neuronal somata supporting a radiant outgrowth of naked axons into the periphery of the culture dish. After the addition of CNS cells, ODCs were identified in these cultures by immunostaining for galactocerebroside (galC), a distinctive glycolipid carried on their surfaces⁸. In some experiments the ODCs were tested for cytoplasmic immunoreactivity using rabbit or goat antisera to myelin basic protein (MBP)⁹. Cell counts indicated that 50-70% of the viable CNS cells adhering to the neuronal culture 24 h after plating were ODCs.

The proliferation of galC⁺ cells was demonstrated by exposing the cultures for 24 h to ³H-thymidine (0.1 $\mu\text{Ci ml}^{-1}$ at 6.7 Ci

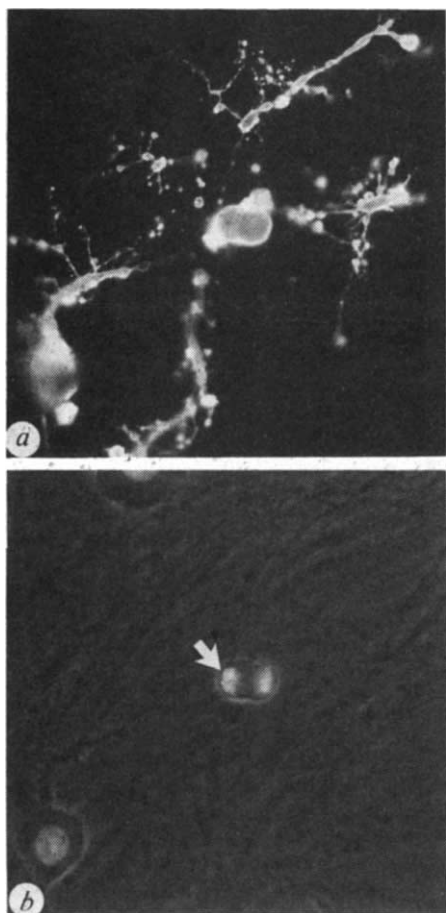


Fig. 2 Mitotic adult galC⁺ ODC in a neurone-containing culture. The cultures were prepared as described in Fig. 1 legend. *a*, Fluorescence image of a dividing cell in a culture fixed 8 days after the addition of spinal cord cells (as described for Fig. 1) to the neuronal cultures. The culture was stained as described for Fig. 1 except that undiluted hybridoma supernatant was used as the primary anti-galC antibody. *b*, Combined phase-contrast and fluorescence image of the cell shown in *a*. Chromosomes are revealed through the enhanced fluorescence of bound Hoechst dye 33342 which was included in the mounting medium (arrow). Note that this mitotic cell has retained some of its processes while dividing. $\times 415$.

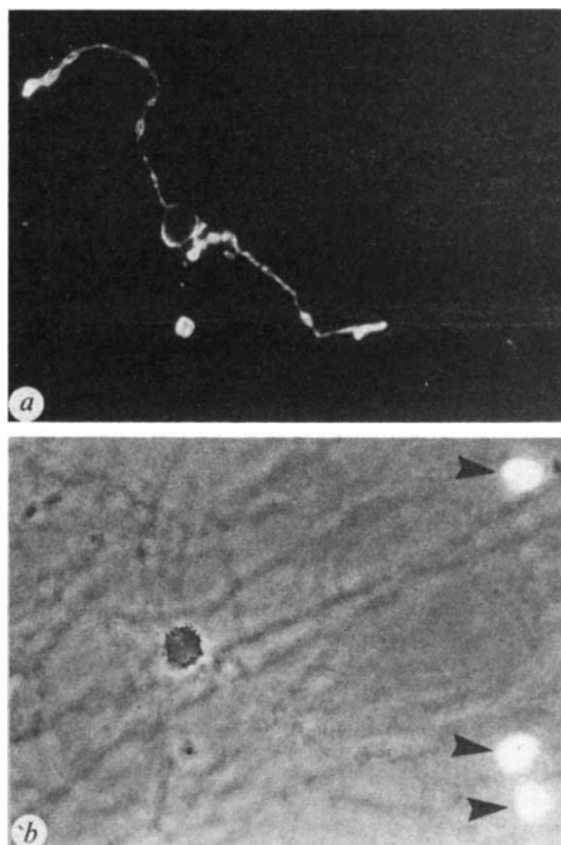


Fig. 3 Thymidine-incorporating adult ODC immunostained with monoclonal anti-galC before exposure to ³H-thymidine. The cultures were prepared as described in Fig. 1 legend. *a*, Fluorescence image of a cell in a culture 8 days after the ODC and neurones were combined (as described for Fig. 1). This culture was incubated with monoclonal anti-galC hybridoma supernatant for 30 min at room temperature, the antibody was washed out and the culture was refed with medium containing 1 μ Ci ml⁻¹ ³H-thymidine (6.7 Ci mmol⁻¹). The culture was incubated for 2 h at 35 °C, then rinsed and incubated with rhodamine-conjugated goat anti-mouse IgG. Note the retention of processes in this cell, which is either in S phase or within 2 h of completing S. *b*, Combined phase-contrast and fluorescence image focused on the plane of the emulsion to illustrate silver grains. The underlying neurites can be seen dimly and out of focus. Three unlabelled nuclei (that is, without superimposed silver grains) revealed by the fluorescence of Hoechst dye 33342 are indicated by arrows. $\times 415$.

in Figs 1–3, the dividing ODCs were either in contact with or very close to bundles of axons.

To further test the requirement for the presence of neurones for ODC division, the mitotic behaviour of ODCs in cultures of neurones was compared with that of ODCs in parallel cultures without neurones (Table 1*a*). In a second series of experiments glial cells were plated onto collagen-coated dishes, and neurones were added to these cultures and allowed to generate axons into the region of the culture containing ODCs (Table 1*b*). In this case, the labelling index of ODCs in the vicinity of axons was compared with that of ODCs in the same culture but in an area not invaded by axons. In both these experimental paradigms ODC proliferation was observed in the presence of axons but not in their absence.

Previously reported results^{2,3} have led to the suggestion that mature ODCs in adult animals retain the ability to divide. Although our results confirm that galC-synthesizing ODCs may retain that ability, we have observed in these experiments that more than 90% of the dividing galC⁺ ODCs exhibit a lack of, or very weak, cytoplasmic immunoreactivity with anti-MBP

antisera compared with that observed in many of the non-dividing ODCs in the same cultures. In addition, we observed that in any given culture the number of colonies of proliferating ODCs was only 3–5% of the total number of galC⁺ ODCs adhering to neurites 24 h after their addition, and that neurone-associated single (apparently non-dividing) galC⁺, MBP⁺ ODCs were numerous even 14 days after their introduction to the neurones. These observations suggest that most of the dividing ODCs may have been recruited from a population of ODCs that were galC⁺ but not actively engaged in myelin synthesis before their isolation from the intact tissue. Although the existence *in vivo* of such a pool of non-myelinating or partially differentiated ODCs has never been demonstrated, it has also not been possible to show that every ODC maintains a connection to a myelin sheath. Furthermore, the recent observation that post-mitotic nonmyelinating Schwann cells expressed galC in peripheral nervous tissues¹⁰ raises the question of whether a parallel may exist in the CNS. While we recognize that our failure to detect significant levels of MBP in dividing ODCs could have other explanations than that suggested above, such

Table 1 Effect of neurones on ^3H -thymidine incorporation in adult oligodendrocytes

Expt	Type of culture	Labelling index (all cells)*	Labelling index (galC ⁺ ODCs)
a	1 DRGN with added glia	28 (22, 34)	41 (37, 44)
	Glia only	22 (17, 27)	0
	2 DRGN with added glia	37 (26, 48)	49 (48, 50)
	Glia only	66 (63, 68)	0
b	Glia with added DRG		
	Cells on axons	58 (56, 59)	41 (40, 41)
	Cells distant from axons	28 (25, 31)	0

The dissociated dorsal root ganglion neuronal (DRGN) cultures were prepared as described in Fig. 1 legend. In *a* the labelling index (defined as the % of cells with more than 10 silver grains over their nucleus) of ODCs adhering to neurones (DRGN+glia) is compared with that of ODCs adhering to parallel empty collagen-coated dishes (glia). The cultures were given $0.1 \mu\text{Ci ml}^{-1}$ ^3H -thymidine on the 7th (expt 1) or 8th (expt 2) day after addition of glia and were immunostained for galC, fixed and processed for autoradiography 24 h later. In *b* the labelling index of ODCs adhering to empty collagen-coated dishes which were in an area invaded by axons growing from an added DRGN network is compared with the labelling index of ODCs in areas of the same culture distant from the invading axons. These cultures were given ^3H -thymidine on the 7th day after the addition of the DRGN network. The values shown are averages of counts of 200–300 cells on each of duplicate cultures; the values in parentheses are those obtained for each culture.

low levels of MBP in the dividing ODCs are not consistent with the suggestion that myelinating ODCs can be induced to divide. We would argue, therefore, that the ability of mature, myelinating ODCs to re-enter the mitotic cycle has not been demonstrated in our own or other experiments^{2,3}.

The data reported here provide the first direct evidence that neurones produce a molecule that is mitogenic for ODCs. We believe that this mitogen may have a significant role in inducing the division of ODCs in adult animals, thereby promoting remyelination after demyelinating events which spare axons. The nature of this mitogen and the manner of its presentation to ODCs remain unknown. It has been amply documented that axons carry on their surfaces a molecule that is mitogenic for the myelin-forming cells of the peripheral nervous system, the Schwann cell^{11–13}. There is recent evidence that the mitogenic molecule is, or depends on, a neuronal surface heparan sulphate proteoglycan¹⁴. Experiments are under way in our laboratory to determine whether neurone-mediated ODC proliferation may also depend on a neuronal heparan sulphate proteoglycan.

We thank Mrs Artree James for technical assistance, Ms Susan Mantia for preparing the manuscript, Dr Barbara Ranscht for monoclonal antibodies to galactocerebroside and for hybridoma cells producing this antibody, Dr John Trotter for goat anti-MBP antiserum, and Dr James Salzer for rabbit anti-MBP antiserum. This work was supported by National Multiple Sclerosis Society grant RG1118.

Binding of angiotensin and atrial natriuretic peptide in brain of hypertensive rats

Juan M. Saavedra, Fernando M. A. Correa,
Laura M. Plunkett, Anita Israel,
Masaki Kurihara & Kazuto Shigematsu

Unit of Preclinical Neuropharmacology, Section on Clinical Pharmacology, Laboratory of Clinical Science, National Institute of Mental Health, NIH, Bethesda, Maryland 20892, USA

Atrial natriuretic peptides, produced in the mammalian cardiac atrium, are released into the general circulation and may be actively involved in the control of blood pressure and in fluid homeostasis as antagonists of the peripheral angiotensin system^{1–4}. Certain cardiovascular effects of atrial natriuretic peptides may be centrally mediated, as binding sites for atrial natriuretic factor (8–33) (ANF) have been localized to the subfornical organ⁵. This circumventricular structure lacks a blood–brain barrier and is therefore accessible to circulating peptides. It contains large numbers of angiotensin II (AII) binding sites, and has been suggested as the main central site of action for circulating AII in the regulation of blood pressure and fluid metabolism^{6,7}. Here we have studied binding sites for rat atrial natriuretic peptide(6–33) (rANP) and AII in the brains of spontaneously (genetic) hypertensive rats (SHR) and their normotensive controls, Wistar Kyoto (WKY) rats⁸, by quantitative autoradiography^{5,9–11}. Binding sites for both peptides were highly localized in the subfornical organ. The number of rANP binding sites was decreased in the subfornical organ of both young (4 weeks old) and adult (14 weeks old) SHR compared with age-matched normotensive controls. Conversely, the number of AII binding sites was higher in both young and adult SHR compared with WKY rats. Our results suggest a central role for rANP and AII in genetic hypertension; they may act as mutual antagonists in brain areas involved in control of blood pressure and fluid regulation.

rANP closely resembles rat atrial natriuretic factor(8–33) containing only two more amino acids (serine and leucine) at the N-terminus¹². Atrial natriuretic peptides have natriuretic and hypotensive effects in normal rats^{1,2} and antihypertensive actions in experimentally hypertensive rats¹³ and in SHR¹⁴; they may lower hypertension in SHR through a direct vasodilator effect, as hypertension in this model is associated with increased peripheral resistance¹⁵.

We have compared AII and rANP binding in brain of young (4 weeks old) and adult (14 weeks old) SHR with that in their age-matched normotensive WKY controls, to determine whether any differences observed occurred early in the development of the hypertension, or were a consequence of the chronic increase in blood pressure. Age-dependent alterations in brain neurotransmitters have been described previously in SHR^{16,17}.

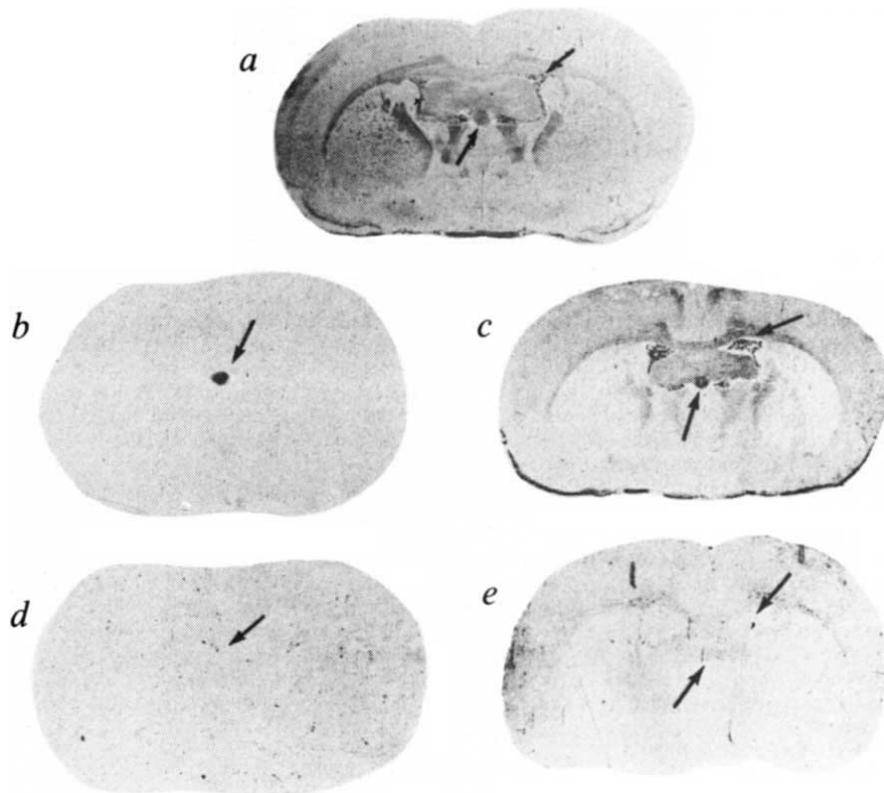
Saturable, single-class binding sites for ^{125}I -rANP were localized in the subfornical organ (Table 1; Figs 1, 2). Nonspecific binding, determined by the addition of unlabelled rANP (Table 1, Fig. 1) or unlabelled ANF(8–33) (results not shown) was less than 30% of the total binding. Scatchard analysis, performed using consecutive sections from individual brains, demonstrated decreased maximum binding capacity (B_{max}) in the subfornical organ of both young and adult SHR compared with age-matched WKY rats (Table 1, Fig. 2). Also, the affinity constant for binding (K_{a}) was lower in the subfornical organ of adult SHR than in that of WKY rats (Table 1).

The choroid plexus also showed a high concentration of ^{125}I -rANP binding sites. Values of B_{max} were lower in choroid plexus of SHR than in age-matched WKY rats (74 ± 13 and 21 ± 3 fmol per mg protein for 4-week-old WKY rats and SHR,

Received 20 September; accepted 21 October 1985.

- Ludwin, S. in *Demyelinating Disease; Basic and Clinical Electrophysiology* (eds Waxman, S. G. & Ritchie, J. M.) 123–168 (Raven, New York, 1981).
- Arenella, L. & Herndon, R. M. *Archs Neurol.* **41**, 1162–1165 (1984).
- Ludwin, S. *Lab. Invest.* **52**, 29–30 (1985).
- Wood, P., Szuchet, S., Williams, A. K., Bunge, B. P. & Aranson, B. G. W. *Trans. Am. Soc. Neurochem.* **14**, 212 (1983).
- Wood, P. & Bunge, R. P. *Nature* **256**, 662–664 (1975).
- Salzer, J. & Bunge, R. P. *J. Cell. Biol.* **84**, 739–752 (1980).
- Wood, P. & Williams, A. K. *Dev. Brain Res.* **12**, 225–241 (1984).
- Ranscht, B., Clapshaw, P. A., Price, J., Noble, M. & Seifert, W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2709–2713 (1982).
- Bologna-Sandru, L. *et al. Brain Res.* **210**, 217–229 (1981).
- Jessen, K. R., Morgan, L., Brammer, M. & Mirsby, R. *J. Cell Biol.* **101**, 1135–1143 (1985).
- Salzer, J., Bunge, R. P. & Glaser, L. *J. Cell Biol.* **84**, 767–774 (1980).
- Cassel, D., Wood, P. M., Bunge, R. P. & Glaser, L. *J. cell. Biochem.* **18**, 433–446 (1982).
- DeVries, G. H., Minier, L. N. & Lewis, B. L. *Dev. Brain Res.* **9**, 87–93 (1983).
- Ratner, N., Bunge, R. P. & Glaser, L. *J. Cell Biol.* **101**, 744–754 (1985).

Fig. 1 Autoradiographic images of ^{125}I -rANP and ^{125}I -[Sar¹]AII binding in rat brain at the subfornical organ level. *a*, Section stained with Luxol fast blue. *b*, Section incubated with 1 nM ^{125}I -[Sar¹]AII and exposed to ^3H -Ultrafilm for 2 days. *c*, Section incubated as in *b*, with the addition of 1 μM unlabelled AII. *d*, Section incubated with 400 pM ^{125}I -rANP and exposed to ^3H -Ultrafilm for 3 days. *e*, Section incubated as in *d*, with the addition of 1 μM unlabelled rANP. All sections are from a 14-week-old WKY rat. Arrows indicate the subfornical organ (*a-e*) and the choroid plexus (*a*, *d* and *e*).



respectively; 120 ± 14 and 61 ± 10 fmol per mg protein for 14-week-old WKY rats and SHR, respectively; $P < 0.05$) but the K_a values showed no difference at either age between the two strains (9.1 ± 1.9 and $11.3 \pm 2.2 \times 10^9 \text{ M}^{-1}$ for 4-week-old WKY rats and SHR, and 7.6 ± 2.0 and $6.3 \pm 0.8 \times 10^9 \text{ M}^{-1}$ for 14-week-old WKY rats and SHR, respectively). No differences in concentration or affinity of rANP binding sites between WKY rats and SHR were found in another forebrain area rich in rANP sites, the olfactory bulb (not shown).

Differences in the numbers of AII binding sites in SHR compared with WKY rats were evident in the subfornical organ of both young and adult animals, but these differences were the reverse of those found for rANP binding sites (Table 1, Fig. 2). SHR exhibited significantly more binding sites for AII than normotensive rats at both ages. Affinity constants, however, were lower in young and adult SHR than in WKY rats (Table 1, Fig. 2). These differences were specific to the subfornical organ, as they were not evident in another circumventricular structure, the area postrema¹⁸.

The subfornical organ has a crucial role in the central effects of circulating AII in the regulation of blood pressure and fluid homeostasis, including regulation of vasopressin secretion¹⁹. Our results suggest that this structure may represent the site for central actions of atrial natriuretic peptides also⁵. The subfornical organ sends projections to the hypothalamus, specifically to the anteroventral/third ventricle (AV3V) region, an area critical for the development and maintenance of experimental hypertension as well as fluid and electrolyte balance²⁰. This area contains the largest accumulation of ANF-positive cells in the brain²¹ and large numbers of AII-positive nerve terminals²² and receptors⁹⁻¹¹. AII nerve cells and terminals²² and the angiotensin-converting enzyme kinase II are also concentrated in the subfornical organ^{23,24}. Thus, the subfornical organ and its hypothalamic projections may represent the site of interactions between circulating and brain neuropeptides such as atrial natriuretic peptides and AII.

Several lines of evidence link both AII and atrial natriuretic peptides with blood pressure regulation. A generalized decrease in the activity of the peripheral AII system²⁵⁻²⁸ and increased

activity of the central AII system occur in SHR²⁹⁻³¹; these animals also have a decreased cardiac ANF content, and increased blood levels of the peptide^{1,32}. Both AII and atrial natriuretic peptides play a part in the secretion of vasopressin, a peptide whose levels are increased in spontaneous hypertension³³. Further, AII stimulates vasopressin release via central mechanisms³⁴. Administration of AII increases blood pressure^{6,20} while injection of ANF has a hypotensive effect^{14,35,36}. This evidence, when considered in conjunction with the present results, suggests that, in comparison with WKY rats, the increased number of AII binding sites and the decreased number of rANP binding sites in the subfornical organ may both be related to increased vasopressin activity in SHR³⁵.

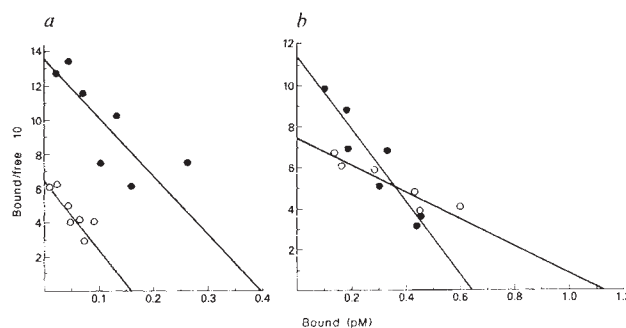


Fig. 2 Scatchard analysis of specific ^{125}I -rANP (*a*) and ^{125}I -[Sar¹]AII (*b*) binding in subfornical organ of 4-week-old SHR (O) and WKY rats (●). Consecutive tissue sections from individual rats were incubated with different concentrations of ^{125}I -labelled peptides as described in Table 1. The data represent the results of typical experiments which were replicated six to seven times per group (see Table 1). Values of K_a , B_{max} and correlation coefficient, r , were as follows. [Sar¹]AII binding in WKY rats: $K_a = 0.88 \times 10^9 \text{ M}^{-1}$, $B_{\text{max}} = 355$ fmol per mg protein, $r = 0.828$; in SHR: $K_a = 0.32 \times 10^9 \text{ M}^{-1}$, $B_{\text{max}} = 702$ fmol per mg protein, $r = 0.940$. rANP binding in WKY rats: $K_a = 3.33 \times 10^9 \text{ M}^{-1}$, $B_{\text{max}} = 113$ fmol per mg protein, $r = 0.742$; in SHR: $K_a = 4.02 \times 10^9 \text{ M}^{-1}$, $B_{\text{max}} = 46$ fmol per mg protein, $r = 0.900$.

Table 1 rANP and AII binding sites in subfornical organ of spontaneously hypertensive rats compared with normotensive controls (WKY)

Peptide	Age (weeks)	Binding capacity, $B_{\max}$ (fmol per mg protein)		Affinity constant, $K_a (\times 10^9 \text{ M}^{-1})$	
		WK	SHR	WK	SHR
Atrial natriuretic peptide	4	85 $\pm$ 12	38 $\pm$ 4*	8.52 $\pm$ 1.30	6.92 $\pm$ 1.53
	14	124 $\pm$ 11	73 $\pm$ 10*	11.10 $\pm$ 1.67	4.25 $\pm$ 0.55*
Angiotensin II	4	444 $\pm$ 85	735 $\pm$ 91*	0.69 $\pm$ 0.10	0.32 $\pm$ 0.06*
	14	848 $\pm$ 136	1,390 $\pm$ 64*	1.90 $\pm$ 0.36	0.89 $\pm$ 0.05*

Groups of six 4-week-old and 14-week-old male SHR and WKY rats (Taconic Farms, Germantown, New York) were housed at a constant temperature with illumination from 06.00 to 18.00 h and given free access to food and water. Blood pressures, measured the day before killing of the animals by an indirect tail-cuff method using a Programmed Sphingomanometer (Narco Biosystems, Inc., Houston, Texas), were 100 ± 8 and 125 ± 10 mm Hg for 4-week-old WKY rats and SHR, respectively, and 120 ± 10 and 184 ± 12 mm Hg for 14-week-old WKY rats and SHR, respectively ($P < 0.01$). The rats were killed by decapitation between 09.00 and 11.00 h and their brains were immediately removed and frozen by immersion in isopentane (-30°C). Within 24 h of killing, tissue sections ($16 \mu\text{m}$) were cut in a cryostat at -14°C , thaw-mounted onto subbed glass slides, and placed under vacuum at 4°C until incubation. rANP binding sites were labelled *in vitro* by incubation with ^{125}I -labelled 3-iodotyrosyl²⁸ rANP (specific activity $1,750 \text{ Ci mmol}^{-1}$; Amersham). Tissue sections were preincubated for 15 min at 20°C in 50 mM Tris-HCl buffer, pH 7.4, then incubated at room temperature for 60 min in 50 mM Tris-HCl buffer, pH 7.4, containing 100 mM NaCl, 5 mM MgCl_2 , 0.5% bovine serum albumin (BSA), $40 \mu\text{g ml}^{-1}$ bacitracin, $4 \mu\text{g ml}^{-1}$ leupeptin, $2 \mu\text{g ml}^{-1}$ chymostatin, $0.5 \mu\text{g ml}^{-1}$ phenylmethylsulphonyl fluoride (PMSF) and ^{125}I -rANP in concentrations ranging from 10 to 400 pM. Nonspecific binding was determined in consecutive sections in the presence of 25 nM to 1 μM of unlabelled rANP (rat atrial peptide, 28 amino acids; Peninsula Laboratories, Inc., Belmont, California). After incubation, the slides were washed three times (for 2 min each) in ice-cold Tris-HCl buffer at 4°C and dried under a cold stream of air. To determine AII binding, tissue sections were preincubated for 15 min in 10 mM sodium phosphate buffer (pH 7.4) containing 120 mM NaCl, 5 mM Na_2EDTA , 0.1 mM bacitracin (Sigma) and 0.2% BSA. The tissue sections were then incubated for 60 min in fresh buffer, with ^{125}I -[Sar¹]AII (specific activity $1,666 \text{ Ci mmol}^{-1}$) in concentrations from 80 pM to 5 nM. Nonspecific binding was determined in alternate sections incubated in the presence of 80 nM to 5.0 μM unlabelled AII (Sigma). After incubation, the slides were washed four times in ice-cold 50 mM Tris-HCl buffer and dried under a stream of air. Autoradiographic images were produced by placing the slides in X-ray cassettes and exposing them to ^3H -Ultrafilm for 4 days. rANP and AII binding sites were quantified by autoradiography with ^3H -Ultrafilm (LKB), followed by computerized microdensitometry and comparison with ^{125}I standards^{10,11}. Scatchard plots were calculated by linear regression. Values are the mean ($\pm$ s.e.m.). Statistical differences between groups were analysed using the Student's *t*-test.

* Significant difference compared with WKY rats ($P < 0.05$).

Independently, an increased number of AII binding sites in the subfornical organ further substantiates the idea of an overactive central AII system in spontaneous hypertension^{29-31,37}, while the decreased number of rANP binding sites suggests that decreased central atrial natriuretic peptide activity may be associated with spontaneous hypertension. As the changes in AII and rANP binding are already present in early hypertensive, young SHR, they are probably not the result of chronic alterations in blood pressure. In the periphery, atrial natriuretic peptides could act as physiological antagonists of the renin-angiotensin system, as they antagonize the increased aldosterone production and vasoconstriction produced by AII¹⁻³. Our results indicate that this antagonism could exist in selected brain areas also.

Received 16 January; accepted 28 February 1986.

- Cantin, M. & Genest, J. *Endocr. Rev.* **6**, 107-127 (1985).
- de Bold, A. J., Borenstein, H. B., Veress, A. T. & Sonnenberg, H. *Life Sci.* **28**, 89-94 (1981).
- Kleinert, H. D. *et al. Hypertension* **6** (Suppl. 1), 1-143-1-147 (1984).
- Campbell, W. B., Currie, M. G. & Needleman, P. *Circulation Res.* **57**, 113-118 (1985).
- Quirion, R. *et al. Peptides* **5**, 1167-1172 (1984).
- Simpson, J. B. *Neuroendocrinology* **32**, 248-256 (1981).
- Mangiapane, M. L. & Simpson, J. B. *Am. J. Physiol.* **239**, R382-R389 (1980).
- Okamoto, K. & Aoki, K. *Jap. Circulation J.* **27**, 282-293 (1963).
- Mendelsohn, F. A. O., Quirion, R., Saavedra, J. M., Aguilera, G. & Catt, K. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1575-1579 (1984).
- Israel, A., Correa, F. M. A., Niwa, M. & Saavedra, J. M. *Brain Res.* **322**, 341-345 (1984).
- Israel, A., Plunkett, L. M. & Saavedra, J. M. *Cell molec. Neurobiol.* **5**, 211-222 (1985).
- Flynn, T. G., de Bold, M. L. & de Bold, A. J. *Biochem. biophys. Res. Commun.* **117**, 859-865 (1983).
- Garcia, R. *et al. Proc. Soc. exp. Biol. Med.* **178**, 155-159 (1985).
- Garcia, R. *et al. Proc. Soc. exp. Biol. Med.* **179**, 396-401 (1985).
- Folkow, B., Hallback, M., Lungen, Y. & Weiss, L. *Acta physiol. Scand.* **79**, 373-378 (1970).
- Saavedra, J. M., Grobecker, H. & Axelrod, J. *Circulation Res.* **42**, 529-534 (1978).
- Saavedra, J. M., Grobecker, H. & Axelrod, J. *Science* **191**, 483-484 (1976).
- Plunkett, L. M. & Saavedra, J. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7721-7724 (1985).
- Mangiapane, M. L., Thrasher, T. N., Keil, L. C., Simpson, J. B. & Ganong, W. F. *Brain Res. Bull.* **13**, 43-47 (1984).
- Hartle, D. K. & Brody, M. J. *Circulation Res.* **54**, 355-366 (1984).
- Jacobowitz, D. M., Skofitsch, G., Keiser, H. R., Eskay, R. L. & Zamir, N. *Neuroendocrinology* **40**, 92-94 (1985).
- Lind, R. W., Swanson, L. W. & Ganten, D. *Neuroendocrinology* **40**, 2-24 (1985).
- Saavedra, J. M. & Chevillard, C. *Neurosci. Lett.* **29**, 123-127 (1982).
- Correa, F. M. A., Plunkett, L. M., Saavedra, J. M. & Hichens, M. *Brain Res.* **347**, 192-195 (1985).
- Chevillard, C. & Saavedra, J. M. *Regul. Peptides* **5**, 333-341 (1983).
- Shioni, K. & Sokabe, H. *Am. J. Physiol.* **231**, 1295-1299 (1976).
- Polsky-Cynkin, R., Reichlin, S. & Fauberg, B. L. *Proc. Soc. exp. Biol. Med.* **164**, 242-247 (1980).
- Niwa, M., Israel, A. & Saavedra, J. M. *Eur. J. Pharmacol.* **110**, 133-136 (1985).

- Stamler, J. F., Brody, M. J. & Phillips, M. I. *Brain Res.* **186**, 499-503 (1980).
- Weyhenmeyer, J. A. & Phillips, M. I. *Hypertension* **4**, 514-523 (1982).
- Phillips, M. I. *Neuroendocrinology* **25**, 354-377 (1978).
- Sonnenberg, H. & Samson, W. K. *Neuroendocrinology* **40**, 277-279 (1985).
- Crofton, J. T., Share, L., Shade, R. E., Allen, C. & Tarnowski, D. *Am. J. Physiol.* **235**, H361-H366 (1978).
- Samson, W. K. *Neuroendocrinology* **40**, 277-279 (1985).
- Seymour, A. A., Marsh, E. A., Mazack, E. K., Stabilito, I. I. & Baine, E. H. *Hypertension* **7** (Suppl. 1), 1-35-1-42 (1985).
- Milosevic, S., Chong, C. K. & Veress, A. T. *Hypertension* **5**, 672-672 (1983).
- Ganten, D., Hermann, K., Bayer, C., Unger, T. & Lang, R. E. *Science* **221**, 869-871 (1983).

Constitutive *c-myc* oncogene expression blocks mouse erythroleukaemia cell differentiation but not commitment

Joseph A. Coppola & Michael D. Cole

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

Mouse erythroleukaemia cells (also called Friend cells) can be isolated from the spleen of certain strains of mice that have been infected with the Friend virus complex¹. The cells resemble pro-erythroblasts and, when exposed to dimethyl sulphoxide (DMSO) or a variety of other chemicals, can be induced to undergo a programme of differentiation which closely resembles the final stages of normal erythropoiesis². This includes the cessation of proliferation and large increases in the production of messenger RNA for both α - and β -globin. In addition, DMSO induces a rapid (<2 h) decrease in *c-myc* mRNA levels^{3,4}. The *c-myc* oncogene is expressed in the majority of proliferating normal cells⁵ and altered expression of the gene has been implicated in the genesis of a wide variety of tumours⁶⁻⁹. To study the influence of oncogene activation on differentiation, we have transfected viral-promoter-driven *c-myc* genes into mouse erythroleukaemia cells. Constitutive *c-myc* expression was found to block DMSO-induced differentiation.

The mouse erythroleukaemia (MEL) cell line F4-12B2 is TK⁻ (lacks thymidine kinase) and can be transfected with high

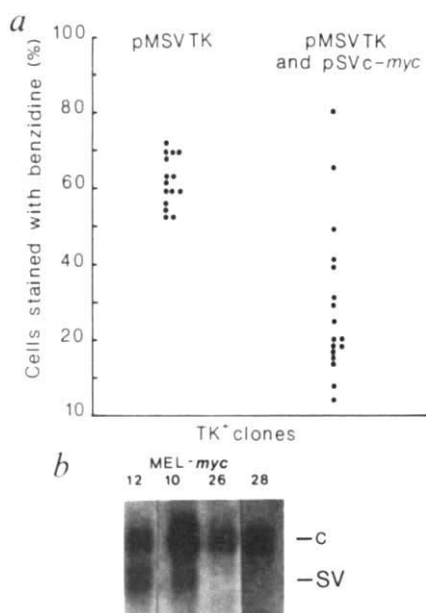


Fig. 1 Differentiation of MEL cells after transfection with a constitutively expressed *c-myc* oncogene. **a**, Percentage of haemoglobin-producing cells. Fifteen cell lines transfected with pMSVtk alone and 17 lines co-transfected with pMSVtk and pSVc-myc were tested for the percentage of cells staining positive with benzidine after 6 days in 2% DMSO. Each dot represents the value obtained for one of the lines. The percentage of cells staining positive for the untransfected F4-12B2 line under these conditions was 62%. **b**, RNAs from four of the pSVc-myc-transfected lines that gave different responses to DMSO treatment were analysed by the Northern blot technique and probed with a *c-myc* probe. The band labelled *c* represents the endogenous 2.5-kilobase (kb) band; the band labelled SV represents the shorter, pSVc-myc-derived band. Clones 12 and 10 were blocked in differentiation, giving only 4% and 8% benzidine-positive cells, respectively. Clones 26 and 28 differentiated normally, giving 48% and 65% benzidine-positive cells, respectively. Quantitation of the relative *c-myc* RNA levels in these lines by densitometry indicated that the ratio of the exogenous to endogenous transcripts was 1:1 in SVmyc 12 and 1:2 in SVmyc 10.

Methods. Cells were grown in Ham's F12 media containing 10% fetal calf serum supplemented with 2 mM glutamine, 100 U ml⁻¹ penicillin, and 100 µg ml⁻¹ streptomycin. HAT concentrate was from Boehringer-Mannheim and was used as suggested by the manufacturer. Transfections were done as in ref. 32. For **a**, 10⁵ cells were plated on 60-mm tissue culture dishes and grown in media (as described in Fig. 1) containing 2% DMSO for 6 days. The medium was changed after day 3. The cells were then stained for haemoglobin as in ref. 34. At least 500 cells per plate were counted and the numbers shown are the average values obtained from at least three separate experiments per cell line. For **b**, RNAs were prepared and analysed by Northern blot as in ref. 26; 2 µg of poly(A)-selected RNA was loaded per lane. The probe was a gel-purified 0.9-kb *Xba*I-*Sst*I fragment containing the mouse *c-myc* second exon.

efficiency by the calcium phosphate precipitation method¹⁰. We co-transfected pSVc-myc, a plasmid containing the *c-myc* protein coding exons under the control of the simian virus 40 (SV40) early promoter^{11,12}, and pMSVtk (ref. 13), a plasmid containing the herpes simplex virus *tk* gene driven by the Moloney sarcoma virus long terminal repeat, into F4-12B2 cells. TK⁺ colonies were selected with hypoxanthine/aminopterin/thymidine (HAT) containing medium. Cells transfected with pMSVtk alone served as controls to show that transfection and conversion of F4-12B2 cells to a TK⁺ phenotype does not affect their capacity to differentiate in response to DMSO.

A number of different clones from pSVc-myc-transfected and control cell lines were induced with DMSO and stained with benzidine after 6 days. Figure 1a shows that the percentage of

cells staining positive for haemoglobin was comparable in the control lines and the untransfected parental cell line. In contrast, the percentage of haemoglobin-producing cells varied considerably among the pSVc-myc-transfected lines. Some lines stained at the level of the controls, others gave very few (<10%) benzidine-positive cells, and still others gave intermediate results. We analysed four of the pSVmyc-transfected lines for the presence of transcripts from the exogenous genes. The lines chosen were two showing a low percentage of benzidine-positive cells after treatment with DMSO (SVmyc 12 and SVmyc 10, 4% and 8%, respectively) and two lines showing a high percentage of benzidine-positive cells after treatment with DMSO (SVmyc 26 and SVmyc 28, 48% and 65%, respectively).

Since pSVc-myc lacks the first, non-coding exon of the *c-myc* gene, transcripts originating from the SV40 promoter of this plasmid and extending to the *c-myc* polyadenylation site are about 300 bases shorter than transcripts from the endogenous *c-myc* gene. When RNA from the pSVc-myc-transfected cell lines was analysed by the Northern blotting technique, SVmyc 12 and SVmyc 10 were found to be expressing transcripts from the exogenous genes at a level 0.5 times that of the endogenous gene. Conversely, SVmyc 26 and SVmyc 28, which differentiated normally, did not express the transfected gene (Fig. 1b). The results of Fig. 1 demonstrate that the expression of a viral-promoter-driven *c-myc* gene blocks the DMSO-induced differentiation of MEL cells. The expression of the exogenous *c-myc* gene induced no other altered growth properties; the cells had the same morphology and grew at the same rate as the controls. It is also important to note that the expression of the exogenous gene does not lead to down-regulation of the endogenous *c-myc* gene, consistent with our earlier results for fibroblasts¹¹.

When MEL cells are treated with DMSO, *c-myc* RNA levels are dramatically reduced within 2 h^{3,4}; this is followed by a return to normal levels within the next 24 h. Since this modulation of *c-myc* RNA levels is one of the earliest measurable responses to induction, it was of interest to determine whether the same effect of DMSO on *c-myc* expression occurred when poorly differentiating lines of pSVc-myc-transfected cells were treated with DMSO. RNA was prepared at various times after DMSO induction from *myc*-transfected and control *tk*-transfected cell lines and analysed for *myc* expression by Northern blot (Fig. 2a,b). The control line showed the previously reported rapid drop in *c-myc* expression, followed by a return to the pre-induction level³. Perhaps surprisingly, the pSVc-myc-transfected line shows virtually the same modulation of endogenous *c-myc* expression as the control line, while the level of exogenous transcripts remains about the same at each point. Other pSVc-myc-transfected lines were found to respond similarly (not shown). This result shows that at least one response to DMSO, modulation of endogenous *c-myc* expression, occurs in the pSVc-myc-transfected cells even though the cells do not show the final differentiated phenotype in response to induction. In addition, it suggests that there is a crucial 'window' (probably from about 1 to 24 h after treatment with DMSO) during which *c-myc* expression must be reduced if the cells are to differentiate fully.

DMSO-induced differentiation of MEL cells leads to a dramatic increase in the transcription of the globin genes¹⁴. To further characterize the *myc*-induced block in differentiation, RNA was prepared at various times after exposure to DMSO from one of the control lines and one of the pSVc-myc-transfected lines which showed a low (<10%) percentage of benzidine-positive cells after induction. Northern blot analysis of this RNA using a β -globin probe is shown in Fig 2c. As expected, given the benzidine staining results, the pSVc-myc-transfected line showed only a small increase in β -globin mRNA level after induction, while globin expression in the control line was induced at least 20-fold. The small increase in the β -globin RNA level in the *myc*-treated line is consistent with the low,

but measurable level of benzidine staining. Thus, the *myc*-induced block in differentiation includes a block in the expression of the β -globin gene, presumably at the level of transcription.

Concentrations of DMSO which induce differentiation of MEL cells also inhibit growth¹. Although cells possessing the ability to grow in DMSO can be isolated from most MEL lines, they arise only at a frequency of 10^{-5} to 10^{-6} (ref. 15). To analyse the influence of exogenous pSVc-*myc* expression on MEL cell growth, cell numbers were counted at regular intervals for pSVc-*myc*-transfected and control MEL cells in the presence and absence of DMSO (Fig. 3). As found previously, DMSO has a strong inhibitory effect on the growth of the control cells. In contrast, DMSO has a negligible effect on the growth of the pSVc-*myc*-transfected cells; in fact several lines have been grown continuously in 2% DMSO for up to 3 months with no significant ill effects.

When MEL cells are treated with DMSO, there is a lag period of from 10 to 24 h before the effects of DMSO become irreversible. Cells transferred to inducer-free medium after exposure to DMSO for less time than this lag period will not withdraw from the cell cycle, synthesize α - and β -globin, or show other changes indicative of terminal differentiation. Cells exposed to DMSO for 24 h or longer will go on to differentiate even if transferred to inducer-free medium¹⁶; these cells are said to be committed and the changes which are presumed to occur while the cells are exposed to inducer during the lag period are termed commitment.

Three pSVc-*myc*-transfected MEL lines were grown continuously in 2% DMSO (>1 month) and showed a low level of benzidine-staining cells. These levels are similar to the levels found when cells are treated with DMSO and stained after only 6 days. In order to test the possibility that these cells are nevertheless committed (that is, that the events which exposure to DMSO triggers have already occurred), we treated the cells with cycloheximide ($10 \mu\text{g ml}^{-1}$) for 30 min. Since the *c-myc* protein has a very short half-life (15–20 min)^{17,18}, this treatment would be expected to drastically reduce *c-myc* protein levels, at least transiently. Of course, this treatment would also affect the level of any other proteins with a rapid rate of turnover. After cycloheximide treatment, the cells were incubated in medium free of both cycloheximide and DMSO for 24 h and then stained with benzidine. The results are shown in Table 1.

Table 1 Differentiation of SVmyc MEL cells growing in DMSO is induced by cycloheximide

Cell line	Growth in DMSO	% Benzidine staining	
		-CX	+CX
SVmyc 12	-	<1%	<1%
SVmyc 12	+	4%	40%
SVmyc 10	-	<1%	<1%
SVmyc 10	+	3%	20%
SVmyc 31	-	<1%	<1%
SVmyc 31	+	11%	53%
MEL	-	<1%	<1%
MSVtk7	-	<1%	<1%

SVmyc-transfected MEL cell lines were grown in standard medium or medium containing 2% DMSO for 30 days, treated for 30 min with $10 \mu\text{g ml}^{-1}$ cycloheximide (CX), and then changed to DMSO-free medium. Cells were stained with benzidine after 24 h. Control lines without exogenous SVmyc were treated with cycloheximide in the absence of DMSO. MEL is the parental line and MSVtk7 is a TK⁺ derivative. SVmyc 31 is a MEL cell line which has been transfected with pSVc-*myc* and its response to DMSO is similar to SVmyc 12 and SVmyc 10.

Each of the lines showed a 5–10-fold increase in the percentage of benzidine-positive cells. The same treatment of non-transfected MEL cells and two control lines did not lead to an increase

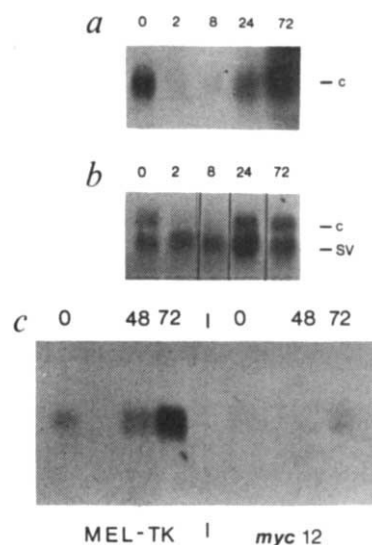


Fig. 2 Expression of endogenous and exogenous *c-myc* in control and pSVc-*myc*-transfected cell lines after treatment with 2% DMSO. *a*, Samples of 20 μg of cytoplasmic RNA were prepared at various times after treatment with 2% DMSO from a pSVtk-transfected cell line and analysed by Northern blot with a *c-myc*-specific probe. The numbers above the lanes refer to the number of hours after DMSO treatment and the band labelled *c* represents the endogenous *c-myc* transcript. *b*, As in *a*, except that 2 μg of poly(A)-selected RNA from SVmyc 12, a pSVc-*myc*-transfected cell line, was used in each lane. The band labelled SV represents the transcript from the exogenous *c-myc* genes. *c*, Production of β -globin mRNA by a control and a pSVc-*myc*-transfected MEL cell line before and after induction with DMSO. The three lanes on the left of the figure show 5 μg of total RNA prepared at 0, 48 and 72 h after exposure to 2% DMSO from a cell line transfected with pSVtk. The RNA was analysed by Northern blot using a β -globin-specific probe. The three lanes on the right show the corresponding experiment done with RNA from SVmyc 12, one of the pSVc-*myc* transfected lines.

Methods. RNA was prepared and analysed as in ref. 26. For *a* and *b*, the probe was a nick-translated 0.9-kb *Xba*I-*Sst*I fragment containing the mouse *c-myc* second exon. For *c*, the probe was a nick-translated 3.35-kb *Eco*RI-*Xba*I fragment containing the mouse β major globin gene, provided by D. Luse.

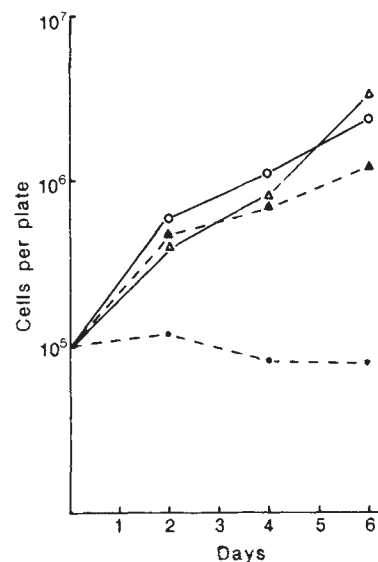


Fig. 3 Growth of SVmyc 12 and a control line with and without DMSO. Cells were plated at 10^5 cells per 60-mm tissue culture dish on day 0. At the indicated times, cells were trypsinized off the plates, centrifuged, resuspended and counted in a haemocytometer. $\blacktriangle$, Control line without DMSO; $\bullet$, control line in 2% DMSO; $\triangle$, SVmyc 12 cells without DMSO; $\circ$, SVmyc 12 cells in 2% DMSO.

in benzidine-positive cells, a finding consistent with previous results¹⁹. Thus, it appears that DMSO-induced commitment events had occurred before cycloheximide treatment; pSVc-myc-transfected MEL cells that had been exposed to DMSO were able to differentiate even in the absence of DMSO when the cycloheximide was removed. The finding that the expression of the endogenous *c-myc* gene is still modulated immediately after DMSO treatment of pSVc-myc-transfected cells (Fig. 2b) is consistent with other aspects of MEL cell commitment also taking place.

Our results argue that the rapid drop in *c-myc* expression observed when MEL cells are treated with inducers is a vital part of the events leading to differentiation rather than a by-product of the process of induction. Because the drop in *c-myc* expression is necessary for differentiation to occur and because it occurs during the time when commitment events are taking place, we propose that modulation of *c-myc* expression is one of the molecular events underlying commitment.

pSVc-myc-transfected cells, which had been grown in DMSO, differentiated when treated briefly with cycloheximide and removed from the inducer. Since commitment is defined as the ability of cells to complete a differentiation programme even in the absence of inducer²⁰, those cells which go on to differentiate after cycloheximide treatment (in the absence of DMSO) were committed. One can therefore infer that some of the changes associated with MEL cell differentiation have occurred even in the presence of constitutive *c-myc* expression. If the half-life of the *c-myc* protein in MEL cells is the same as its half-life in other systems (15–20 min)^{17,18}, cycloheximide treatment of pSVc-myc-transfected MEL cells which had been grown in DMSO would be expected to dramatically reduce *c-myc* protein

levels. Of course, the levels of other short-lived proteins would also be reduced. Therefore, one must be cautious in interpreting the increased differentiation after cycloheximide treatment as being due solely to a transient reduction of *c-myc* protein levels.

We thank Eileen Kuempel and Don Luse for providing the MEL cells and helpful advice, and Steve McKnight and Robert Weinberg for pMSVtk and pSVc-myc plasmids. We also thank members of our laboratory, especially Steve Piccoli and Ameeta Kelekar. J.A.C. was supported by a Damon Runyon-Walter Winchell Cancer Fund Fellowship, DRG-813. M.D.C. is the recipient of a Faculty Research Award from the ACS. This work was also supported by research grants from the ACS and the NIH.

Received 27 November 1985; accepted 10 March 1986.

1. Marks, P. A. & Rifkind, R. A. *Rev. Biochem.* **47**, 419–448 (1978).
2. Friend, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **68**, 378–382 (1971).
3. Lachman, H. M. & Skoultschi, A. I. *Nature* **310**, 592–594 (1984).
4. Lachman, H. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5323–5327 (1985).
5. Gonda, T. J. *et al. Molec. cell. Biol.* **2**, 617–624 (1982).
6. Shen-Ong, G. L. C. *et al. Cell* **31**, 443–452 (1982).
7. Taub, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7837–7841 (1982).
8. Little, C. D. *et al. Nature* **306**, 194–196 (1983).
9. Corcoran, L. M. *et al. Cell* **37**, 113–122 (1984).
10. Spandidos, D. A. & Paul, J. *EMBO J.* **1**, 15–20 (1982).
11. Keath, E. J. *et al. Cell* **39**, 339–348 (1984).
12. Land, H. *et al. Nature* **304**, 596–602 (1983).
13. Graves, B. J. *et al. Molec. cell. Biol.* **5**, 1948–1958 (1985).
14. Ganguly, S. & Skoultschi, A. I. *J. biol. Chem.* **260**, 12167–12173 (1985).
15. Harrison, P. R. *et al. Cell* **14**, 61–70 (1978).
16. Gusella, J. *et al. Cell* **9**, 221–229 (1976).
17. Rabbitts, P. H. *et al. EMBO J.* **4**, 2009–2015 (1985).
18. Hann, S. R. & Eisenman, R. N. *Molec. cell. Biol.* **4**, 2486–2497 (1984).
19. Levenson, R. & Housman, D. *J. Cell Biol.* **82**, 715–725 (1979).
20. Levenson, R. & Housman, D. *Cell* **25**, 5–6 (1981).

A human homoeo box gene specifically expressed in spinal cord during embryonic development

A. Simeone*, F. Mavilio†, L. Bottero†, A. Giampaolo‡, G. Russo‡, A. Faiella*, E. Boncinelli* & C. Peschle†

* International Institute of Genetics and Biophysics, CNR, Viale Marconi, 10, 80125 Naples, Italy

† Department of Haematology, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy

‡ Division of Obstetrics and Gynecology, Ospedale Generale di Avellino, 83100 Avellino, Italy

Several genes involved in the determination of *Drosophila* body segments share a conserved DNA sequence of 183 nucleotides termed the homoeo box¹. Homologous homoeo box sequences have been detected in the genome of species ranging from insects and anellids to vertebrates^{2,3}, and a number of homoeo box-containing genomic DNA clones have been isolated from *Xenopus*, mouse and human^{4–11}. We have recently isolated human complementary DNA clones containing homoeo box sequences, representing transcripts from four different genes¹². We report here the nucleotide sequence of one of these clones (HHO.c10) and show that the corresponding gene is transcribed in human embryos and fetuses at 5–10 weeks post-conception. A major polyadenylated transcript of ~2.1 kilobases (kb), as well as RNA species of higher relative molecular mass (M_r), are specifically expressed at a constant level in spinal cord throughout this developmental period.

The nucleotide sequence of clone HHO.c10 is shown in Fig. 1, along with the conceptual translation of its longest open reading frame. The 5' terminus of this clone encodes the 7th amino-acid residue of the homoeo domain, whereas the carboxy terminus of the predicted protein product lies 15 amino-acid residues

downstream from the end of the domain. A putative 961-base-pair (bp) 3'-untranslated region (3'-UT) contains a polyadenylation signal (AAUAAA) 21 nucleotides upstream from the poly(A) tail. The homoeo domain encoded in HHO.c10 is identical to those of human *Hu-1* (ref. 11) and mouse *H24.1* (ref. 10) genomic clones. Homologies with other homoeo domains are shown in Fig. 1 legend. At the level of DNA sequence, the HHO.c10 homoeo box is identical to that of the *Hu-1* clone^{11,13}. This identity extends throughout the reported sequence of *Hu-1*¹³ up to nucleotide 240 of HHO.c10 (Fig. 1), except for two silent nucleotide differences in the putative translated sequence and three others downstream from the termination codon. Therefore, the HHO.c10 clone is derived from a transcript of a human gene identical or allelic to *Hu-1*.

Previous studies have shown that genes containing homoeo box regions are expressed in ontogenic development in insects, amphibians and mammals^{2,4,5,10,13,14}. To investigate homoeo box gene expression in early human development, we used the HHO.c10 cDNA insert as a probe in Northern blot analysis of total or poly(A)⁺ RNA from embryonic and fetal tissues. Human embryos and fetuses were obtained virtually intact from legal curettage abortions at 5–10 weeks from conception (see also ref. 15). Age of fertilization of 5–8-week-old embryos was carefully established by morphological staging, according to multiple precise criteria¹⁶. Fetuses of 9–10 weeks were staged on the basis of standard age/crown-rump length plots¹⁷. The integrity of the examined embryos and fetuses allowed a dating error of as little as ± 2 days¹⁶.

Northern blot analysis of poly(A)⁺ RNA obtained from whole embryos at 6, 7 and 8 weeks indicated that HHO.c10 represents a gene expressed in human embryogenesis: a major transcript of ~2.1 kb is detected together with a minor band of ~5.6 kb (Fig. 2a). The 2.1-kb band is present in the dissected central nervous system (CNS, that is, brain plus spinal cord) from 5–6-week-old embryos, but not in 7–8-week-old brain and limb buds, nor in limbs and liver from 6–8-week-old embryos (Fig. 2b).

```

30                                60
ala tyr thr arg tyr gln thr leu glu leu glu lys glu phe his phe asn arg tyr leu
GCG TAT ACC CGC TAC CAG ACC CTG GAG CTG GAA AAG GAG TTC CAC TTC AAC CGC TAC CTG

90                                120
thr arg arg arg arg ile glu ile ala his ala leu cys leu ser glu arg gln ile lys
ACC CGG CGA CGG CGC ATC GAG ATC GCC CAC GCA CTC TGC CTG TCC GAG CGC CAG ATC AAG

150                                180
ile trp phe gln asn arg arg met lys trp lys lys asp asn lys leu lys ser met ser
ATC TGG TTC CAG AAC CGG CGC ATG AAG TGG AAG AAG GAC AAC AAA TTG AAA AGT ATG AGC

210                                240
leu ala thr ala gly ser ala phe gln pro trm
CTT GCT ACA GCT GGC AGC GCC TTC CAG CCC TGAGCCGCCAGGAGGCCAGCGGCCAAGAGCCCGT

270                                300
GCCACCCCGAGCCCGCCCTCCCAATCTCCCGCTCGCGGCCCGCTGGGACCGGTCCACAGCCTGCCTC

330                                360                                390
GCCTTGTTGTACGATATTTCGTTTGTATCTAGGCTTCTCTGTGCTCCCTCTCTCATGGATGGTTATCTGTTATTAT
TGTTAATAATAATAATTATTATTATTTCTTCCATGCTCCCAACTCCTTCTGTTGCCAAATCCGCCAGTGTTC

420                                450                                480
TGAATGTTTGTGTGTGTTGAGCTTTCCCCAGGAAAAAAGAAATGCATGTTTAAATGTGAATC

510                                540
TCCCTCCCATCTGTGTTCTTAACCTATTATTAAGAGGATGATCGCTGATTTTGTGTTTGTGCTGGAATCTC

570                                600                                630
TGTAAGGGGCGCAGTTGAGGTGGGAGTAGTCCGAGTGGGTCAGCTGAGCTGCTCGAGATGGAGTCCCTTTTC
ATTCTCTCTCTCTCCCTCTCTCACTCCCTAGGCCAAGTCTCTAGGGGCTTGGTCTTAGGGTGGGAGGGGCTAGGGA

660                                690                                720
GGACCAAGGGATGGTATTGAGAAGAGAGAAAGAGATAGTGAATTAAGTCTCTGCTGGTAGGCCCAACAAG

750                                780
GCCTAGTCTGGGAGTATACGGAACAAAAATGATCCTCAGTGCAAAATGCTTGTGATTCTCTGTGAATCCATCGGGT

810                                840                                870
CTGGCTAGAGGGCCCAAGCTTGTAAATATGGGGATAGTCTGGTCAGACCCATCTCCCTTACCCATCTTGTCTCA
AGACCATTTGTAGTGAGCGAGTGATGCTGTGCTACGTGTGAATCTGTCTTGCGGGCTGTCTCAGTGATTCGCTT

1050                                1080                                1110
TTGGTATTGTTTGTAGCTTCTCGGAAGTCAAAATATGTTTCCCGCACTCC

1140                                1170

```

Fig. 1 Nucleotide sequence of the human cDNA clone HHO.c10, shown together with the conceptual translation of its longest open reading frame (trm, termination codon). This sequence is virtually identical from nucleotide 1 to 240 to that reported for the human *Hu-1* genomic clone¹³, thus indicating that HHO.c10 originates from a transcript of a human gene identical or allelic to *Hu-1*. The homoeo box region, which lacks the first 18 nucleotides in this clone, is underlined. The predicted peptide sequence of the homoeo domain, identical to that of human *Hu-1*¹¹ and murine *H2.1* (= *Mu1*) clone^{10,13}, has the following homologies with other homoeo genes (expressed as number of identical residues out of a total of 61 amino acids): *Antennapedia*, 55 (ref. 3); *Xenopus MM3*, 54 (ref. 5); *Xenopus AC1*, 54 (ref. 4); mouse *Hox-1*, 46 (ref. 6); mouse *m6*, 54 (ref. 7); human *Hu-2*, 51 (ref. 11). The polyadenylation signal preceding the poly(A) tail is also underlined.

Methods. Isolation of human cDNA clones containing homoeo box sequences from simian virus 40 (SV40)-transformed human fibroblasts has been reported previously¹². DNA extraction, restriction endonuclease mapping and subcloning were carried out according to standard procedures²⁷. Appropriate fragments were subcloned in the pEMBL8 plasmid²⁸ and sequenced as described by Maxam and Gilbert²⁹.

To localize precisely the expression of HHO.c10 in the CNS, we dissected separately brain vesicles and spinal cords from embryos and early fetuses ranging from 5 to 10 weeks after conception. Dissection of brain from cord was performed at the level of the pontine flexure separating metencephalon from myelencephalon, the latter remaining attached to the spinal cord. Brains and cords of the same age were pooled, and RNA extracted and analysed by Northern blot hybridization (Figs 2c, 3). Analysis of poly(A)⁺ RNA from 6–8-week spinal cord showed the predominant 2.1- and 5.6-kb bands, as well as at

least three other transcripts of >8, 5.4 and 2.8 kb, respectively (Fig. 2c) (see also below). Conversely, HHO.c10 expression was undetectable in brain RNA, with the exception of a faint 2.1-kb band at 7 weeks (Fig. 2c): further studies are clearly required to clarify the significance of this observation. Virtually no hybridization signal was detected in poly(A)⁺ RNA from a control 6-week embryo on removal of brain vesicles, spinal cord and backbone rudiments (Fig. 2c), thus strengthening the concept of tissue-specific expression of the HHO.c10 homoeo gene in human embryonic development.

Analysis of poly(A)⁺ RNA from spinal cord also showed a faint band in the position of 28S ribosomal RNA (Fig. 2c); rehybridization with a human genomic 28S rRNA probe¹⁸ demonstrated that this band co-migrates with 28S (data not presented), thus indicating that it may be derived from cross-hybridization. On the other hand, rehybridization with a human 18S rRNA genomic probe¹⁸ demonstrated that 18S rRNA migrates slightly but distinctly below the 2.1-kb band (not shown), thus excluding cross-hybridization at the level of the 2.1-kb transcript. Furthermore, competition experiments indicated a marked decrease of the 2.1-kb signals on addition of excess cold HHO.c10 cDNA to the hybridization mixture (results not shown).

To assess whether the bands detected in spinal cord RNA may be derived from cross-hybridization to homoeo box-containing transcripts other than HHO.c10, we rehybridized a filter to a subclone containing part of the putative 3'-UT region of clone HHO.c10. As the relative intensity of both the 2.1-kb and the higher-*M_r* bands was unaffected (Fig. 2c), we conclude that these transcripts are derived from the homoeo gene represented by clone HHO.c10.

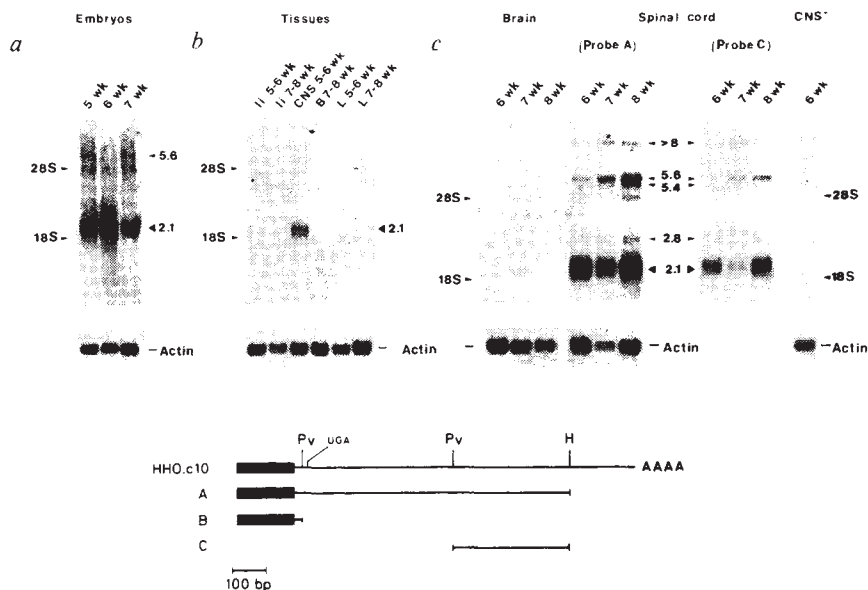
In an attempt to improve the resolution of the broad 2.1-kb band, poly(A)⁺ RNA from a 6-week spinal cord was run extensively on a 1.2% agarose gel and hybridized in turn to the whole cDNA clone (probe A), the homoeo box (B) and the 3'-UT region (C). No finer bands were detected in the 2.1-kb region (actually spanning from 2.2 to 1.8 kb), even after very short exposures, thus suggesting a possible length heterogeneity of this class of transcripts (see also ref. 33). Here again, no significant difference was observed between hybridization signals obtained with the three different subclones.

We further extended our analysis to total RNA from brain and spinal cord at 5–10 weeks from conception (Fig. 3). The 2.1-kb transcript, apparently absent in brain, is relatively abundant at nearly constant level in spinal cord RNA throughout this developmental period. As expected, the specific transcripts of higher *M_r* are also observed, together with the cross-hybridizing 28S rRNA. Taking into account the abundance of the β -actin messenger RNA¹⁹, we estimate that an average of >100 molecules per cell of the 2.1-kb transcript are accumulated in 5–10-week spinal cord. Obviously, the possibility exists that the transcript is heterogeneously expressed in different subregions and/or cell populations of the cord, thus implying an even higher number of transcripts in positive cells.

These observations demonstrate tissue-specific expression of a homoeo gene homologous to the previously reported genomic clone *Hu-1*¹¹ in spinal cord from human 5–8-week embryos and 9–10-week fetuses. A murine homoeo box gene homologous to *Hu-1* is expressed in mouse embryos starting from as early as 7.5 days post-coitum (corresponding to mid-third week in human gestation)^{10,14,20,21}, and is enriched in spinal cord at day 12.5 (6 weeks in humans)¹⁰. This gene is transcribed in a major 2.3-kb transcript in developing mouse embryos¹³. Altogether, these findings demonstrate that a homoeo box-containing gene is actively expressed in mammalian embryos, with a specific localization in the spinal cord and possibly some regions of the developing brain. In the human spinal cord this gene is expressed at a virtually constant level through the 5–10-week developmental and morphogenetic period, in a major heterogeneous class of polyadenylated transcripts averaging 2.1 kb and at least

Fig. 2 *a*, Expression of HHO.c10 in poly(A)⁺ RNA from whole human embryos at 5, 6 and 7 weeks (wk) from conception. 28S and 18S rRNAs are shown as size markers. Sizes are in kb. *b*, Expression of HHO.c10 in poly(A)⁺ RNA from pools of organs or body parts dissected from human embryos at 5–6 or 7–8 weeks: li, limb buds or limbs; CNS, central nervous system (brain vesicles + spinal cord); B, brain; L, liver. *c*, Expression of HHO.c10 in poly(A)⁺ RNA obtained from pools of brains and spinal cords of human embryos at 6–8 weeks from conception. The major transcript of 2.1 kb observed after hybridization to both probes A and C is indicated, together with minor transcripts of higher *M_r*. The last lane (CNS[−]) shows poly(A)⁺ RNA obtained from a 6-week embryo deprived of the CNS (that is, head and spinal cord) and backbone rudiments by microdissection. The map of the entire HHO.c10 human cDNA clone and the three probes derived by subcloning is shown at the bottom. The homoeo box region is indicated by a black box, UGA is the putative transcription termination codon (see Fig. 1). Pv, *Pvu*II; H, *Hind*III.

Methods. Total RNA was extracted from single embryos by the guanidine-HCl or thiocyanate technique³⁰. Poly(A)⁺ RNA was obtained by one passage on oligo(dT)-cellulose columns, run in 2–3-μg aliquots on 1.0% agarose-formaldehyde gels²⁷, transferred to nitrocellulose paper by Northern capillary blot³¹, hybridized to 10⁷ c.p.m. of probe A labelled by nick-translation to a specific activity of 3–6 × 10⁸ d.p.m. per μg, washed at high stringency (final wash: 0.1 × SSC, 0.1% SDS at 65 °C) and exposed for 1–3 days to Kodak XR-5 X-ray films in a X-Omatic intensifying screen cassette. After removal of the first probe, the filter was rehybridized to a chicken β-actin probe³² for normalization.



four higher-*M_r* RNA species. Interestingly, the *Hu-1* genomic clone detects a 1.6-kb transcript in the human teratocarcinoma line NT2/D1 induced to differentiate by retinoic acid¹³; further studies are required to clarify the significance of this difference in transcript length between normal human or murine embryos and *in vitro* differentiating human teratocarcinoma cells.

In *Drosophila* embryos, the ventral nervous system (cord) shows a discontinuous, metameric distribution of transcripts²² and protein products^{23–25} of the *Ultrathorax* (*Ubx*) homoeotic gene complex. The functional significance of the conserved homoeo domains in vertebrates is unknown, and might differ from that in *Drosophila*²⁶. Nevertheless, the observation that a

homoeo box-containing gene is specifically expressed in the neural tube is in line with the hypothesis that these genes are involved in the determination of the segmented body plan in higher organisms.

We thank G. Cossu, L. Lania, G. Grimaldi and P. P. Di Nocera for helpful discussions. This work was supported in part by grants from CNR, Progetti Finalizzati 'Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie' to E.B. and C.P. (grant no. 84.00902.51), 'Oncologia' to E.B., C.P. (84.00730.44) and F.M. (84.00672.44), and 'Medicina Preventiva e Riabilitativa' to C.P. (85.00707.56).

Received 13 December 1985; accepted 4 March 1986.

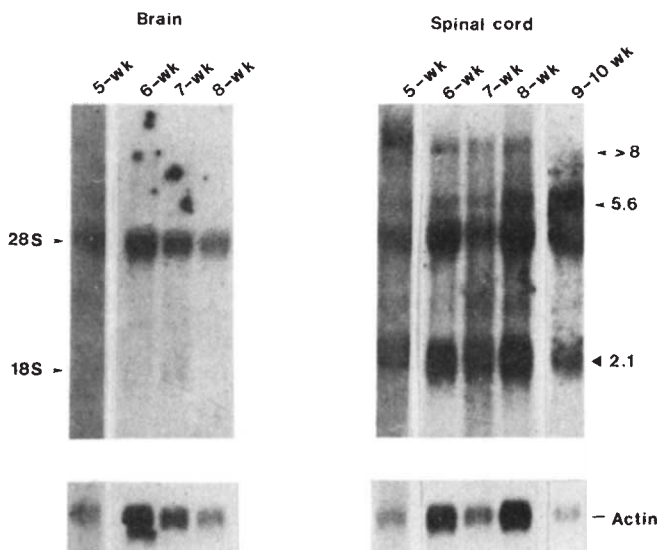


Fig. 3 Expression of HHO.c10 in pools of brains and spinal cords from human embryos at 5–10 weeks from conception. Total cellular RNA (10 μg) was hybridized to probe A (see Fig. 2), and the relative amount of transcripts normalized with a β-actin probe.

- Gehring, W. J. *Cell* **40**, 3–5 (1985).
- McGinnis, W., Levine, M. L., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428–433 (1984).
- McGinnis, W., Garber, A. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 403–408 (1984).
- Carrasco, A. E., McGinnis, W., Gehring, W. J. & De Robertis, E. M. *Cell* **37**, 409–444 (1984).
- Mueller, M. M., Carrasco, A. E. & De Robertis, E. M. *Cell* **39**, 157–162 (1984).
- McGinnis, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. *Cell* **38**, 675–680 (1984).
- Colberg-Poley, A. M., Voss, S. D., Chowdhury, K. & Gruss, P. *Nature* **314**, 713–718 (1985).
- Rabin, M. *et al. Nature* **314**, 175–177 (1985).
- Joyner, A. L. *et al. Nature* **314**, 173–175 (1985).
- Jackson, I. J., Schofield, P. & Hogan, B. *Nature* **317**, 745–748 (1985).
- Levine, M., Rubin, G. M. & Tijan, R. *Cell* **38**, 667–673 (1984).
- Boncinelli, E. *et al. Cold Spring Harb. Symp. quant. Biol.* **50** (in the press).
- Hauser, C. A. *et al. Cell* **43**, 19–28 (1985).
- Colberg-Poley, A. M. *et al. Cell* **43**, 39–45 (1985).
- Peschle, C. *et al. Nature* **313**, 235–238 (1985).
- Moore, K. L. *The Developing Human* 3rd edn (Saunders, Philadelphia, 1982).
- Shepard, T. H., Nelson, T., Dakley, G. P. & Lemire, R. J. in *Monitoring Birth Defect and Environment. The Problem of Surveillance* (eds Hook, E. B., Janerich, D. T. & Porter, I. H.) 29–43 (Academic, New York, 1971).
- La Volpe, A. *et al. J. molec. Biol.* **183**, 213–223 (1985).
- Ponte, P., Gunning, P., Blau, E. & Kedes, L. *Molec. cell Biol.* **3**, 1783–1791 (1983).
- Hart, C. A. *et al. Cell* **43**, 19–28 (1985).
- Joyner, A. L. *et al. Cell* **43**, 29–37 (1985).
- Akam, M. E. & Martinez-Arias, A. *EMBO J.* **4**, 1689–1700 (1985).
- White, R. A. H. & Wilcox, M. *Cell* **39**, 163–171 (1984).
- Beachy, P. A., Helfand, S. L. & Hogness, D. S. *Nature* **313**, 545–551 (1985).
- White, R. A. H. & Wilcox, M. *EMBO J.* **4**, 2035–2043 (1985).
- Hogan, B., Holland, P. & Schofield, P. *Trends Genet.* **1**, 67–74 (1985).
- Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645–1655 (1983).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
- Chirgwin, J. M. *et al. Biochemistry* **18**, 5294–5300 (1979).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
- Cleveland, D. W. *et al. Cell* **20**, 95–105 (1980).
- Hahn, W. E., Chaudari, N., Beck, L., Wilber, K. & Peffley, D. *Cold Spring Harb. Symp. quant. Biol.* **48**, 465–475 (1983).

Homoeo-domain homology in yeast *MAT α 2* is essential for repressor activity

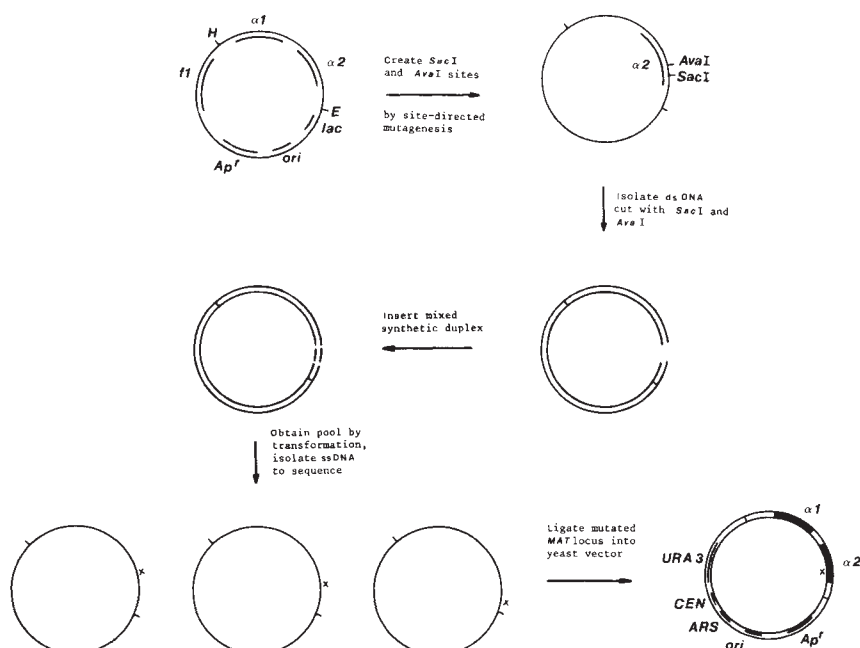
Susan D. Porter & Michael Smith

Department of Biochemistry, Faculty of Medicine,
University of British Columbia, Vancouver, British Columbia,
Canada V6T 1W5

The *MAT α* locus of the yeast *Saccharomyces cerevisiae* encodes two regulatory proteins, $\alpha 1$ and $\alpha 2$, which are responsible for determining the α -cell type¹⁻⁴. *MAT α 1* is a positive regulator of α -cell-type-specific genes, and *MAT α 2* is a negative regulator of α -cell-type-specific genes. *MAT α 2* also determines the a/α diploid cell type, in conjunction with the *MAT α* product, $\alpha 1$, by repressing haploid cell-type-specific genes⁵. The *MAT α 2*-encoded protein binds specifically *in vitro* to a DNA sequence found upstream of several a -specific genes⁶ and is thus thought to exert its control directly at the transcriptional level of target genes. In an initial attempt to understand the molecular basis of the interaction of $\alpha 2$ with DNA, we have saturated with missense mutations the segment of $\alpha 2$ that is weakly homologous to a conserved prokaryote DNA-binding structure⁶⁻⁸ and to a portion of the higher eukaryote homoeo domain^{9,10} to ascertain the possible functional significance of this homology in $\alpha 2$. We report here that most of the amino-acid residues in $\alpha 2$ which correspond to conserved amino acids in the prokaryote DNA-binding proteins and in the homoeo domain are essential for the two repressor activities of $\alpha 2$, that is, the repression of a -specific genes and of haploid-specific genes. Mutations in a subset of these amino-acid residues more severely affect the ability to repress a -specific genes than haploid-specific genes.

Figure 1 shows the strategy for saturation missense mutagenesis of the segment of *MAT α 2* coding for the block of amino acids most highly conserved relative to the homoeo domain. The wild-type 56-base-pair (bp) segment was replaced by a synthetic duplex fragment containing a low level of mutant nucleotides at designated positions. The design of the synthesis was optimized for the production of one missense mutation per molecule (see Fig. 3).

Fig. 1 Outline of strategy for producing random point mutations in the homoeo-domain homology. This strategy is similar to that described elsewhere^{31,32}. All manipulations were performed on the 4-kilobase *HindIII*-*EcoRI* *MAT α* locus inserted into the vector pEMBL8(-) (ref. 33). Creation of the two unique restriction sites, *AvaI* and *SacI*, bordering the mutagenic 'cassette' was accomplished by oligonucleotide-directed mutagenesis³⁴. The plasmid was cleaved with *AvaI* and *SacI* and the intervening fragment was replaced with a heterogeneous synthetic duplex fragment where, at designated sites, a low level of mutant bases had been incorporated. (For sequence details, see Fig. 2.) The ligation mixture, which was first subjected to cleavage with *SacI* to select against reinsertion of the wild-type fragment, was transformed into *E. coli*. A pool of the resultant molecules was obtained and individual clones selected, after sequencing by the dideoxy chain terminator method³⁵, for insertion into the yeast plasmid YCp50. DNA purification and enzymatic steps³⁶ and yeast transformations³⁷ were performed according to standard protocols. *Ap^r*, ampicillin resistance gene. H and E, *HindIII* and *EcoRI* sites, respectively; fl, the region of the fl genome containing the *cis*-acting elements required for DNA replication and morphogenesis.



To determine the ability of the mutants to repress an a -specific gene, they were inserted into a low-copy-number yeast vector and transformed into a yeast strain deleted in the *MAT* locus and containing the *Escherichia coli lacZ* gene inserted in-frame within *STE6* (ref. 11). β -Galactosidase activity in this strain is inhibited by a functional *MAT α 2* gene product. Of the mutants tested, *ste6::lacZ* transcript levels were shown to correlate with β -galactosidase levels (data not shown). The ability of these transformants to mate with a cells was assayed to assess the overall function of the mutant $\alpha 2$ products (Fig. 2).

The ability of the $\alpha 2$ mutants to repress a haploid-specific gene and to determine the a/α cell type was assayed by transformation into a *MAT α* strain containing a *lacZ* fusion to the haploid-specific gene¹² *HO* (constructed by L. Breeden and K. Nasmyth), and by transformation into a *MAT α /MAT*-deletion diploid followed by quantitation of sporulation. Again, β -galactosidase levels of the *ho::lacZ* strain transformants correlated with *ho::lacZ* transcript levels (data not shown). The results of the analysis of some of the mutants are presented in Fig. 3.

One striking observation is that alterations at six of eight positions identical to those of the homoeo domain prevented $\alpha 2$ -mediated repression of a -specific genes. Conversely, all six of the mutations at non-conserved residues gave rise to functional gene products. A different pattern is seen with the analysis of $\alpha 1$ - $\alpha 2$ activity in a/α cells, where mutations in all but two of the conserved residues yield partially or fully active gene products.

Crystallographic, biochemical and genetic analyses of the λ DNA-binding proteins *cI* and *cro*, and *E. coli* catabolite-activating protein (CAP), have resulted in a model in which a common α -helix-turn- α -helix structure is in contact with the major groove of the operator DNA (for a review see ref. 13). The second helix is located within the groove, where it makes contact with specific nucleotides, while the first lies across the second and positions it properly. The *in vitro* construction of mutants within the second helix of the bacteriophage 434 repressor structure which alter the DNA-binding specificity of the protein^{14,15}, and the selection of analogous mutants in the *E. coli* CAP¹⁶ provide further evidence that the second helix of the helix-turn-helix structure is responsible for DNA-binding specificity. Many other DNA-binding proteins have sequence homologies to key residues of this structure, which is suggestive

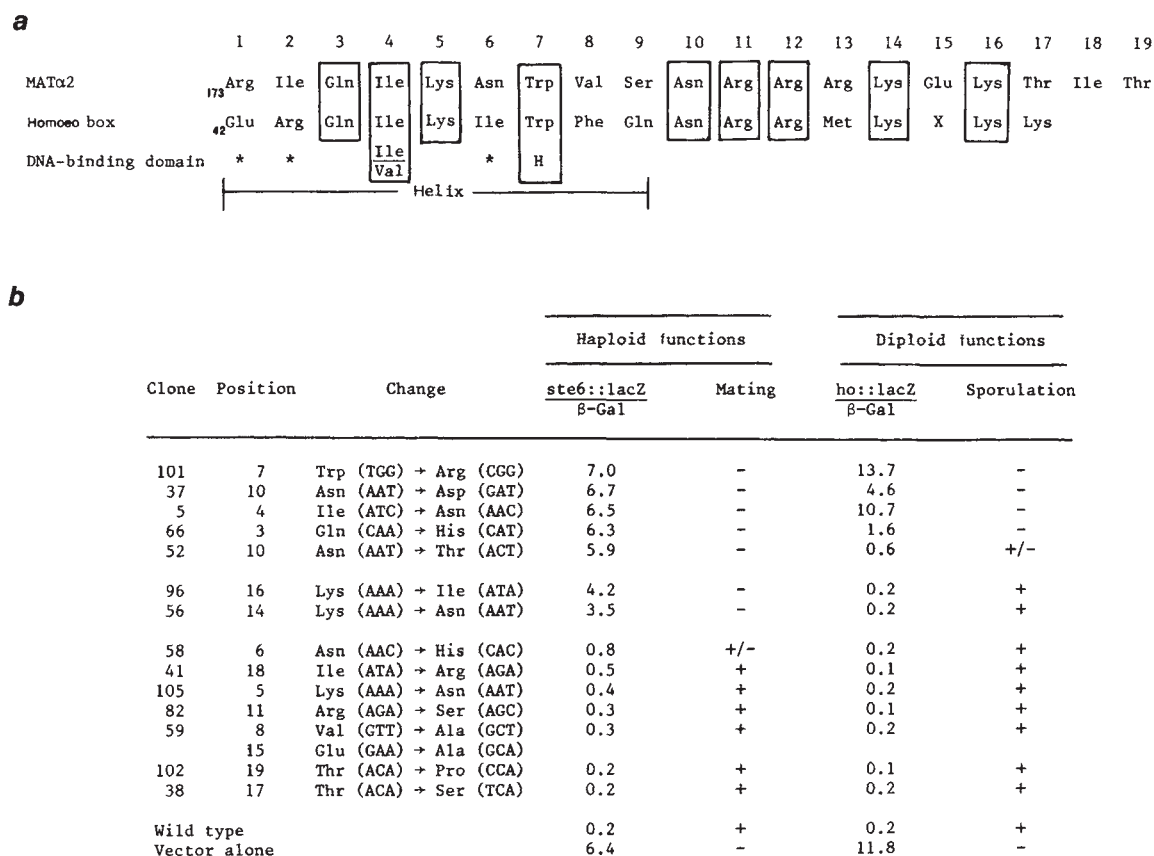


Fig. 2 MAT α 2 mutations occurring in the region homologous to the homoeo domain, and their phenotypes. **a** Aligns a portion of MAT α 2 sequence with the carboxy-terminus of the homoeo-domain sequence and the second helix of the prokaryote DNA-binding domain consensus⁸. Conserved residues are boxed. X denotes a position in the homoeo-domain consensus sequence that is not conserved. H and * in the DNA-binding domain sequence refer respectively to a hydrophobic residue and those residues thought to be in contact with DNA³⁸. Changes at residue positions indicated by the numbering system in **a** and corresponding phenotypes are shown in **b**.

Methods. MAT α 2 mutations within the MAT α locus were inserted into YCp50 and transformed into either SP2 (*mat::CAN ste6::lacZ his3 his4 leu2 trp1 ura3*), which was derived from a cross between strains MH52-3C (a strain constructed by M. Hall using a *mat::CAN* locus constructed *in vitro* by P. Siliciano and K. Tatchell and consisting of a replacement by the *CAN* gene of the entire DNA sequence within the MAT locus) and 1160 (containing *ste6::lacZ*, ref. 11), K1107 (*HMRa MATa hmla ho::lacZ his3 leu2 trp1 ura3 ade2 can1 met*⁻), constructed by L. Breeden and K. Nasmyth, or SP3 (*MATa/mat::CAN ho::lacZ/ho HIS3/his3 HIS4/his4 trp1/trp1 leu2/leu2 ura3/ura3*), a diploid formed by mating MH52-3C and 1402 (ref. 39). MH52-3C, 1160 and 1402 were provided by I. Herskowitz, and K1107 by K. Nasmyth via I. Herskowitz. β -Galactosidase assays of SP2 and K1107 transformants were performed as described in ref. 40. Units were measured as $1,000 \times A_{420}/A_{600} \times \text{time (min)} \times \text{volume (ml)}$. Values were quite variable between individual transformants (up to 40%), and an average of two separate transformants is indicated. Quantitative mating analyses were performed as follows: 2×10^6 cells of each transformant and of D311-3A (*MATa lys2 his1 trp2*) were mixed and collected on filters, which were placed on YPD plates for 6 h at 30 °C. Cells were resuspended in water, vortexed and appropriate dilutions spread on plates selecting for transformant, tester or diploids. Sporulation analyses of diploid transformants were carried out after growth on sporulation plates⁴¹ for 3 days at 30 °C. For both mating and sporulation analyses, - refers to values <1% of control values, +/- refers to values 20–80% of control values, and + refers to values >90% of those of the wild type.

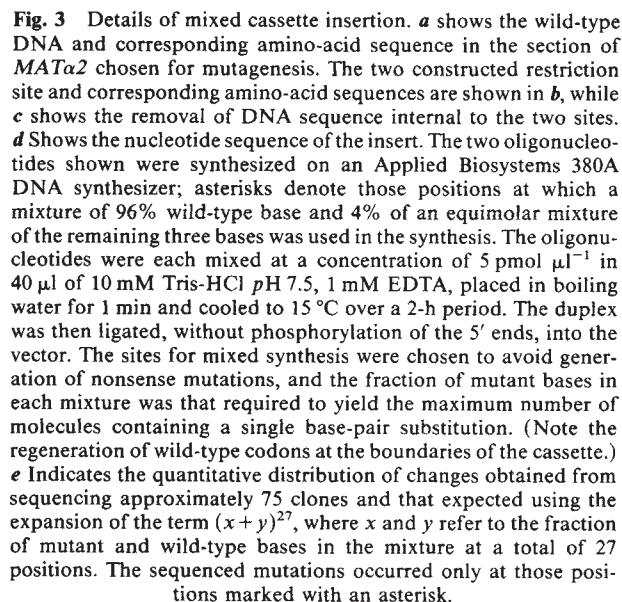
Sporulation percentages refer to the fraction of asci with respect to the total number of cells.

of a common tertiary structure^{6,7}. The conserved residues among at least 11 regulatory proteins occur at the interface of the two helices, in the turn, and at the internal hydrophobic residues not exposed to solvent. Because of the general conservation of these residues in the homoeo domain, it was postulated to have a similar DNA-binding role in the proteins that contain it⁸. Results presented in Fig. 3 are entirely consistent with this relationship for MAT α 2.

Mutation of the conserved Ile/Val 4 (which, in the helix-turn-helix structure, is a point of contact between the two helices) to Asn, and the conserved hydrophobic nature of residue 7 (a residue buried within the protein) to Arg, which would be expected to destabilize the putative helix-turn-helix or surrounding structure, gave rise to two of the most severe mutant phenotypes in both haploid and diploid functions. A substitution at Gln 3 (which, in the equivalent position of the *lac* repressor, drastically reduces DNA binding^{17,18}) is also non-functional. It is interesting that these residues at positions 3, 4 and 7 are

conserved identically (with one Ile 4 to Val substitution¹⁹) in all *Drosophila*, human, mouse and frog homoeo-domain sequences determined to date^{9,19–25}. They are also present (with conservative changes in the hydrophobic residue) in the MAT α 1-encoded protein and in two mating-type regulatory genes of *Schizosaccharomyces pombe*, which share no other homology with the *cerevisiae* genes beyond this region (J. Burke, D. Beach and M.S., unpublished results). Experiments are in progress to determine the ability of these α 2 mutants to bind DNA *in vitro*.

Two mutations (clones 96 and 56) in the basic sequence C-terminal to the proposed DNA-binding domain yield gene products that are fully active in diploid cell function (for example, *HO* repression) yet inactive in haploid cell function (for example, *STE6* repression). These results are interesting in that they suggest a slightly different role for the α 2 polypeptide in each of its two functions. It is possible that the difference in activities of α 2 is due, at least in part, to the probability that



We thank I. Herskowitz, R. Jensen, K. Wilson, K. Nasmyth and M. Hall for providing strains and plasmids; T. Atkinson for the synthesis of oligonucleotides; and D. Goodin, J. Ngsee, B. Gilks and especially I. Herskowitz for comments on the manuscript. The research was supported by the MRC and NCI of Canada. S.D.P. is a recipient of a MRC studentship, and M.S. is a Career Investigator of the MRC.

20. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. *Cell* **40**, 37-43 (1985).
21. Levine, M., Rubin, G. M. & Tjian, R. *Cell* **38**, 667-673 (1984).
22. McGinnis, W. J., Hart, C. P., Gehring, W. J. & Ruddle, F. H. *Cell* **38**, 675-680 (1984).
23. Colberg-Poley, A. M. *et al.* *Cell* **43**, 39-45 (1985).
24. Joyner, A. L., Kornberg, F., Coleman, K. G., Cox, D. R. & Martin, G. R. *Cell* **43**, 29-37 (1985).
25. Carrasco, A. E., McGinnis, W., Gehring, W. J. & DeRobertis, E. M. *Cell* **37**, 409-414 (1984).
26. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. *Nature* **289**, 244-250 (1981).
27. Hall, M. N., Hereford, L. & Herskowitz, I. *Cell* **36**, 1057-1065 (1984).
28. Miller, A. M., MacKay, V. L. & Nasmyth, K. A. *Nature* **314**, 598-603 (1985).
29. MacKay, V. L. & Manney, T. R. *Genetics* **76**, 255-271 (1974).
30. Sprague, G. F., Rine, J. & Herskowitz, I. *Nature* **289**, 250-252 (1981).
31. Wells, J. A., Vasser, M. & Powers, D. B. *Gene* **34**, 315-323 (1985).
32. McNeil, J. B. & Smith, M. *Molec. cell Biol.* **5**, 3545-3551 (1985).
33. Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645-1655 (1983).
34. Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468-500 (1983).
35. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
36. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
37. MacKay, V. L. *Meth. Enzym.* **101**, 325-343 (1983).
38. Ebringt, R. H. in *Protein Structure, Function, and Design* (ed. Oxender, D.) (Liss, New York, in the press).
39. Russell, D. W. *et al.* *Molec. cell Biol.* (in the press).
40. Ruby, S. W., Szostak, J. W. & Murray, A. W. *Meth. Enzym.* **101**, 253-269 (1983).
41. Rothstein, R. J., Esposito, R. E. & Esposito, M. *Genetics* **85**, 35-54 (1977).

PITC derivatives in amino acid analysis

from Steven A. Cohen, Brian A. Bidlingmeyer and Thomas L. Tarvin

Refined methods for separating PTC-amino acids on reverse phase columns may pose a challenge to traditional ion exchange techniques.

FOR over 25 years, amino acid analysis (AAA) has relied on the ion-exchange method of Moore, Spackman and Stein¹ for the separation and quantitation of free amino acids. Detection is routinely accomplished by derivatization with ninhydrin or, more recently, with ortho-phthalaldehyde. In the past decade researchers have employed derivatization of amino acids²⁻⁴ prior to separation by reverse phase liquid chromatography (LC) in order to overcome limitations in speed and sensitivity associated with the ion-exchange procedure. However, these

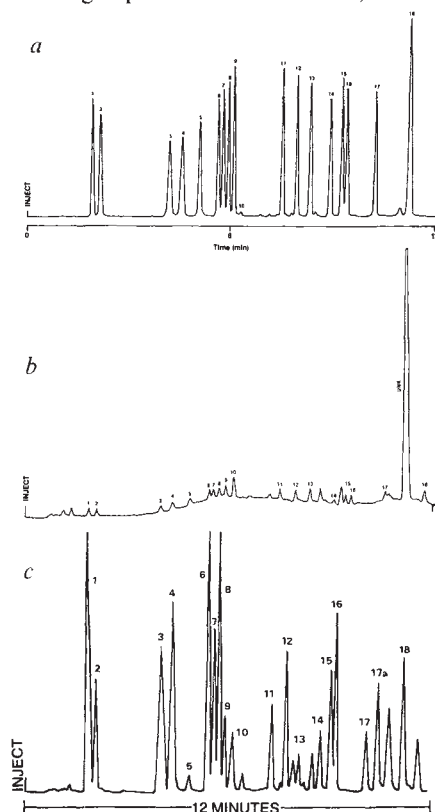


Fig. 1 Chromatography of hydrolysates and hydrolysate standards. Chromatographic conditions: Eluent A = 0.14 M sodium acetate, 0.05 per cent (v/v) TEA, pH 6.40 with 6 per cent (v/v) acetonitrile; eluent 3 = 60 per cent acetonitrile in water. Peak 1 = Asp, 2 = Glu, 3 = Ser, 4 = Gly, 5 = His, 6 = Arg, 7 = Thr, 8 = Ala, 9 = Pro, 10 = phenylthiourea, 11 = Tyr, 12 = Val, 13 = Met, 14 = Cys, 15 = Ile, 16 = Leu, 17 = Phe, 17a = Trp, 18 = Lys. *a*, Analysis of 250 pmol of each PTC-amino standard injected of a total 12.5 nmol each derivatized. *b*, Same as in *a* except 1 pmol was injected out of a total 100 pmol derivatized. *c*, Analysis of 0.5 µg of hen egg white lysozyme out of 5.0 µg hydrolysed with 20 µl 4 M methanesulphonic acid containing 0.1 per cent (w/v) tryptamine.

Table 1 Compositional analyses determined by precolumn derivatization with PITC

	Bovine serum albumin Hydrolysis of			Oxidized insulin B chain Hydrolysis of			Lysozyme MSA Hydrolysis of	
	5.0 µg Found	Published	0.5 µg Found	0.5 µg Found	Published	0.1 µg Found	5.0 µg Found	Published
CA	—		—	2.0	(2)	1.7	—	
Asp	53.3	(53)	49.7	1.0	(1)	1.1	22.1	(21)
Glu	79.3	(78)	73.1	3.0	(3)	2.8	5.6	(5)
Ser	26.6	(28)	27.0	1.1	(1)	1.1	8.6	(10)
Gly	19.2	(15)	16.0	2.8	(3)	2.9	12.5	(12)
His	16.4	(17)	14.7	1.9	(2)	1.5	1.1	(1)
Arg	24.1	(23)	23.5	1.1	(1)	1.2	7.2	(11)
Thr	31.2	(34)	32.4	1.0	(1)	1.1	7.3	(7)
Ala	46		46	2		2	12	
Pro	28.5	(28)	28.9	1.1	(1)	1.1	2.4	(2)
Tyr	19.5	(19)	18.6	1.9	(2)	1.6	3.7	(3)
Val	33.2	(36)	35.4	2.9	(3)	2.4	6.5	(6)
Met	4.0	(4)	4.0	0		0	1.4	(2)
Ile	13.3	(14)	16.0	0		0	5.0	(6)
Leu	62.3	(61)	64.0	4.1	(4)	3.8	8.3	(8)
Phe	28.3	(26)	29.7	3.2	(3)	2.7	3.3	(3)
Trp	ND		ND	ND		ND	5.1	(6)
Lys	64.3	(59)	63.4	1.1	(1)	1.7	6.6	(6)

— Samples were hydrolysed *in vacuo* with 6 M HCl or 4M methanesulphonic acid (lysozyme). Aliquots corresponding to 10–20 per cent of the hydrolysate were injected for quantitation. Compositions were calculated by normalizing the amounts recovered to the known quantity of Ala. CA, cysteic acid; ND, not determined.

techniques have suffered from one or more of the following drawbacks: poor derivative stability, less than quantitative reaction yields, interference from reagent peaks, and a lack of reactivity with secondary amino acids.

Recently we published a complete methodology which overcomes these problems⁵. The method enables high speed, low-level amino acid analysis (AAA) of protein hydrolysates and is based upon a gaseous acid hydrolysis of the proteins followed by the formation of a phenylthiocarbamyl (PTC) derivative of the amino acids via derivatization with phenylisothiocyanate (PITC)⁶. The derivatized amino acids undergo LC separation on a bonded phase column, followed by ultraviolet detection at 254 nm. With a totally volatile reagent system, a large reagent excess can be used to drive derivatization to completion in under 20 minutes without introducing significant interferences. This integrated procedure results in a reproducible, reliable, accurate and extremely rapid (12 minutes) analysis with a detection limit of one picomole for each amino acid. More recently applications of the PTC-derivatization methodology have appeared determining the amino acid composition at low levels in plant material⁷, in growth factors isolated from bovine neural tissue⁸, and in angiogenin (an angiogenic protein) from human carcinoma cells⁹.

With appropriate care in sample preparation, accurate compositional analyses can be obtained with 0.5 µg of protein hydrolysate, and 0.1 µg samples can provide good data for most or all amino acids. However, environmental contamination, possibly from atmospheric dust, can contribute to high values for serine and glycine. Substitution of methanol for ethanol in the derivatizing mixture allows the analysis of samples in the presence of non-volatile buffer salts and detergents¹⁰. In addition, protein tryptophan content can be measured in methanesulphonic acid (MSA) hydrolysed samples via minor modifications in sample preparation¹⁰.

Current research has focused on expanding the capabilities of PTC-AAA to the analysis of food and feed samples, with accuracy and reproducibility comparable to ion-exchange analysis¹¹.

This review summarizes the recent advances in the application of PTC-AAA to different sample types. We also discuss for the first time the analysis of free amino acids such as those present in biological fluids, which may contain 40 or more amino-containing components. High sensitivity inherent to PTC-AAA allows accurate plasma amino acid concentrations to be determined from as little as 25 µl of plasma. The data presented here demonstrate that the reverse phase PTC methodology can be used for a 60-minute analysis of physiological amino acids compared to

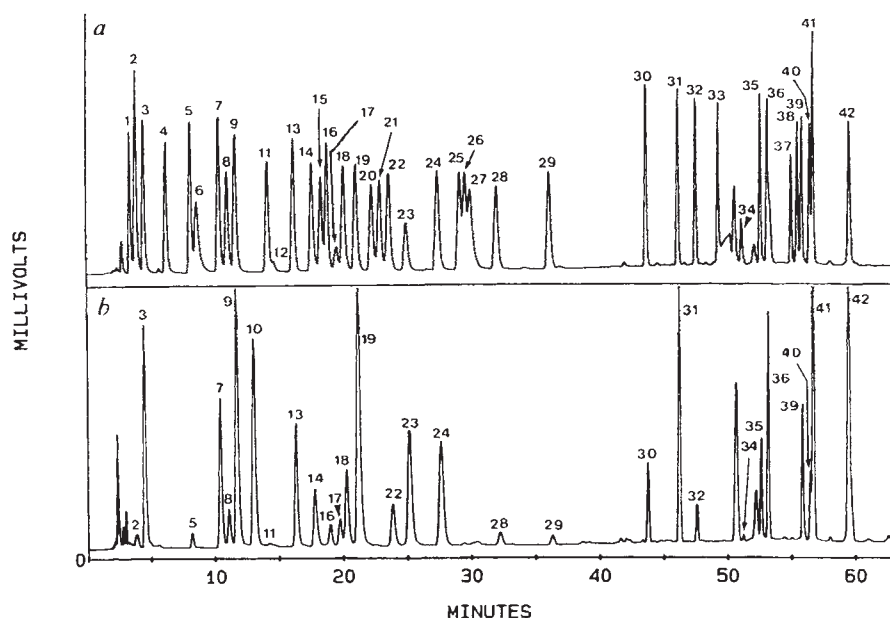


Fig. 2 Chromatography of physiological amino acids. Eluent A = 70 mM sodium acetate, pH 6.55 with 2.5 per cent (v/v) acetonitrile; eluent B = 45 per cent acetonitrile, 15 per cent methanol, 40 per cent water. Identification of peaks: 1 = phosphoserine, 2 = Asp, 3 = Glu, 4 = α -amino adipic acid, 5 = hydroxyproline, 6 = phosphoethanolamine, 7 = Ser, 8 = Asn, 9 = Gly, 10 = Gln, 11 = β -Ala, 12 = sarcosine, 13 = taurine, 14 = His, 15 = γ -aminobutyric acid, 16 = citrulline, 17 = phenylthiourea, 18 = Thr, 19 = Ala, 20 = β -aminoisobutyric acid, 21 = carnosine, 22 = Arg, 23 = methionine sulphone, 24 = Pro, 25 = 1-methylhistidine, 26 = anserine, 27 = 3-methylhistidine, 28 = creatinine, 29 = α -aminobutyric acid, 30 = Tyr, 31 = Val, 32 = Met, 33 = cystathionine, 34 = Cys, 35 = Ile, 36 = Leu, 37 = hydroxylysine, 38 = hydroxylysine, 39 = Phe, 40 = Trp, 41 = ornithine, 42 = Lys. *a*, Analysis of a mixture of standard physiological amino acids. The amount injected was 500 pmol for most of the components. *b*, Analysis of human serum. The amount injected corresponds to 2.5 μ l of serum.

the traditional ion exchange approach, which takes approximately 3–6 hours for completion.

Proteins and peptides

Proteins and peptide samples (0.1–5.0 μ g) were hydrolysed at 110°C for 20–24 h using a vapour phase hydrolysis technique with constant boiling HCl containing 0.5–1.0 per cent phenol (v/v). Following removal of residual HCl under vacuum, samples were neutralized with excess triethylamine (TEA) and derivatized with methanol-TEA-PITC-water (7:1:1:1)^{5,10}. After 20 minutes at room temperature, excess reagent and solvents were removed under reduced pressure and the samples reconstituted in 5 mM phosphate buffer at pH 7.5 prior to injection.

A fast, high resolution separation of amino acids is shown in Fig. 1a. A combination of rapid gradient generation across a high efficiency, application-specified PICO TAG column, a high sensitivity fixed wavelength detector, and eluents chosen to have low ultraviolet background, enables detection of subpicomole amounts (Fig. 1b) of the derivatized amino acids. Derivative stability is excellent with less than 5 per cent loss in 2 days at 4°C and the response is linear over the range 10–2,000 pmol. Reproducibility for retention time and area response are very good: average relative standard devi-

ations are 0.2 and 1.5 per cent, respectively⁵. Minor changes in operating conditions can be used for samples which contain unusual amino acids, such as collagen.

The high sensitivity of the method makes possible accurate compositional analyses on very small (<1 μ g) samples; however, hydrolysis of submicrogram amounts of peptides and proteins requires great care in sample handling to avoid contamination of samples. With appropriate controls, calculated compositions for peptides and proteins were equivalent at the 0.5 and 5.0 μ g levels (Table 1), corresponding to amounts as little as 8 pmol of protein. With hydrolysis of 0.1 μ g accurate compositions (Table 1) can be obtained for most amino acids.

The flexibility of this method has been shown with the analysis of a wide variety of different sample types and matrices. Tryptophan content in peptides and proteins can be quantitated in samples hydrolysed with MSA¹⁰. Hydrolysis was performed batchwise using 20 μ l 4 M MSA containing 0.1 per cent tryptamine (w/v) per sample. After hydrolysis, samples were neutralized with 4 M KOH and derivatized as previously described.

Even though the acid salt is present in high concentration following neutralization, good compositional analyses with typical tryptophan yields of some 60–90 per cent can be obtained (see Table 1).

The resulting chromatogram is shown in Fig. 1c.

Physiological samples

In comparison with hydrolysate-type samples, the analysis of amino acids in biological fluids increases the demands on sample preparation and the chromatographic separation. Ultrafiltration (UF) for protein removal from biological fluids, previously reported in studies on nucleic acids¹², drugs¹³, and other compound classes¹⁴, was employed to avoid interference in derivatization caused by non-volatile acids (for example, sulphosalicylic or perchloric) often used for deproteinization. Physiological samples were deproteinized by UF using hydrophilic 10,000 MW cutoff membranes from Millipore Corporation (Bedford, MA, USA). Volumes from 50 to 250 μ l were diluted 1:1 with an internal standard solution (0.2 mM methionine sulphone in 0.1 M HCl) and centrifuged at 3,000 r.p.m for 10–15 minutes to provide the driving force for filtration. Aliquots from 10–50 μ l were dried and derivatized in the same manner as hydrolysed proteins.

This preparation gives several advantages over conventional techniques for protein removal. First, there is no discernible precipitation with subsequent sampling problems due to flocculent particles in the supernatant. Second, no losses are incurred through weak association with proteins as can occur with precipitation techniques, and third, there are no significant changes in sample volume, as there are with the precipitation procedures.

Validation of physiological amino acid analysis via the PTC-AAA method, including recovery of amino acids from biological fluids and comparison to ion-exchange analysis, is now under way in our laboratory. □

Steven Cohen, Brian Bidlingmeyer and Thomas Tarvin are at Waters Chromatography, Division of Millipore Corp., 34 Maple Street, Milford, Massachusetts 01757, USA.

- Moore, S., Spackman, S.H. & Stein, W.H. *Analyt. Chem.* **30**, 1185–1190 (1958).
- Englehart, H., Asshauer, J., Neue, U. & Weigand, N. *Analyt. Chem.* **46**, 336–340 (1974).
- Hill, D.W., Walters, F.H., Wilson, T.D. & Stuart, J.D. *Analyt. Chem.* **51**, 1338–1341 (1979).
- Haag, A. & Langer, K. *Chromatographia* **7**, 659–662 (1974).
- Bidlingmeyer, B.A., Cohen, S.A. & Tarvin, T.L. *J. Chromatogr.* **336**, 93–104 (1984).
- Koop, D.R., Morgan, E.T., Tarr, G.E. & Coon, M.J. *J. biol. Chem.* **257**, 8472–8480 (1982).
- Budde, R.J.A., Holbrook, G.P. & Chollet, R. *Archs. Biochem. Biophys.* **242**, 293–290 (1985).
- Lobb, R.R. & Fett, J.W. *Biochemistry* **23**, 6295–6299 (1984).
- Strydom, D.J. *et al. Biochemistry* **24**, 5486–5494 (1985).
- Cohen, S.A., Tarvin, T.L. & Bidlingmeyer, B.A. *Proc. American Protein Chemists Society Mtg*, San Diego (in the press).
- Bidlingmeyer, B.A., Cohen, S.A., Tarvin, T.L. & Frost, B. *J. Assn. Off. Analyt. Chem.* (in the press).
- Hartwick, R.A., Van Haverbeke, D., McKeag, M. & Brown, P.R. *J. Liq Chromatogr.* **2**, 725–744 (1979).
- Miller, T.D. & Pinkerton, T.C. *Analytica. chim. Acta* **170**, 295–300 (1985).
- Cook, S.E., Pinkerton, T.C., Sadler, G. & Nelson, P.E. *J. Chromatogr.* **325**, 450–455 (1985).

Analysis in the ion exchange tradition

from Greg Ogden

Amino acid analysis by ion exchange chromatography is just as important an analytical tool in 1986 as it was in the late 1950s, when Moore and Stein first described an automated AAA instrument.

IN contrast to more recently developed techniques for amino acid determination, which rely on reverse phase HPLC separation of derivatized amino acids, the traditional method employs ion exchange chromatography followed by post-column derivatization with ninhydrin or, where higher sensitivity is required, orthophthalaldehyde, for detection. Whilst the basic technique has changed relatively little from that presented in Moore and Stein's original publication¹, significant developments in methodology and instrumentation have left their mark on today's generation of dedicated amino acid analysers.

With ion exchange chromatography, separation of the native amino acids requires minimal sample preparation and in some cases amounts to merely filtering prior to injection. The post-column ninhydrin reaction takes place in a low volume, high temperature reaction coil, where the optimized reaction conditions give full colour development in about one minute. Upon reaction with ninhydrin, most components produce a complex which absorbs strongly at 570 nm; however, the imino acids, proline and hydroxyproline, form a complex with different absorbance characteristics. Therefore a dual channel detector is required, with the second channel set to 440 nm.

From hydrolysate beginnings

Traditional protein and peptide characterization begins with sample hydrolysis in 6 M HCl followed by evaporation or lyophilization and injection in a pH 2.2 buffer^{2,3}. Separations of protein hydrolysates, which typically contain 18 components or less, can then be achieved using a system of three sodium citrate buffers⁴. Careful optimization of chromatography parameters will effect the separation of unusual components such as contaminants in impure samples, artefacts introduced by the protein chemist (for example *S*-carboxymethylcysteine, homoserine and homoserine lactone), or as naturally occurring compounds (for example post-translationally modified amino acids).

The other classical application for amino acid determination is the analysis of free amino acids in physiological fluids such as urine, plasma, serum and cerebrospinal fluid.

Physiological samples are inherently more complex than protein hydrolysates. In contrast to the 18 components found in

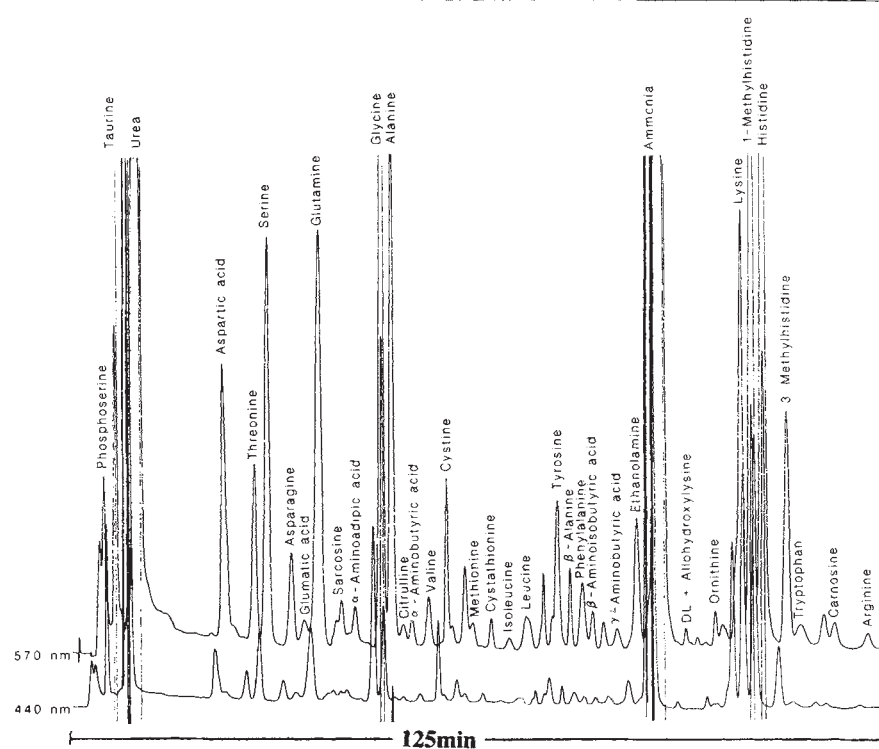


Fig. 1 Chromatogram showing the separation of human urine (40 µl) on an LKB Alpha Plus amino acid analyser using a lithium citrate buffer system.

most protein hydrolysates, physiological fluids can contain between 25 and 50 amino acids and other ninhydrin-positive compounds. Owing to this complexity, a system of five lithium citrate buffers has been developed to effect the separation of physiological fluids and related sample⁶.

Sample preparation for most physiological fluids is limited to deproteinization, usually with 5-sulphosalicylic acid⁷, followed by filtration, although some samples may also require subsequent pH adjustment for optimal separation. Sample preparation procedures have also been published for the analysis of plant and animal tissues^{8,9}. Figure 1 shows an example of a physiological fluid separation obtained using the lithium citrate buffer system.

Future analysis

There is no reason to suppose that current trends towards faster, more sensitive analysis will not continue. Reverse-phase HPLC methods have gained in popularity because they appear to offer the increase in speed and sensitivity which some workers seem to require. However, the traditional, dedicated amino acid analyser can also benefit from advances in HPLC tech-

nology, and the late 1980s will without doubt see a new generation of Moore and Stein machines employing miniaturized components, higher operating pressure and smaller and more efficient ion-exchange resins and microbore columns. These advances, coupled with developments in buffer and reagent systems, will mean that the proven reliability and reproducibility of the established technique will continue to survive as the method of choice for a great many workers. □

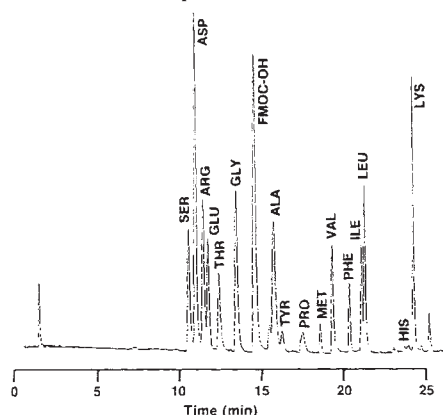
1. Spackman, D.H., Stein, W.H. & Moore, S. (1958) *Analyt. Chem.* **30**, 1190-1206 (1958).
2. Davies, M.G. & Thomas, A.J. *J. Sci. Fd. Agric.* **24**, 1525-1540 (1973).
3. Moore, S. & Stein, W.H. *Methods in Enzymology* **6**, 819-831 (1963).
4. Condon, G.D., Baldar, N.A. & Andrews, R.P. *Protein Chemistry Note No. 30* (1 KR Biochrom Ltd., Cambridge, UK, 1984).
5. Mason, V.C., Bech-Anderson, S. & Rudemo, M. *Proc. 3rd EAAP Symp. Protein Metabolism Nutrition* **1**, 351-355 (1980).
6. Andrews, R.P. & Condon, G.D. *Protein Chemistry Note No. 31* (LKB Biochrom Ltd., Cambridge, UK 1985).
7. Mondino, A., Bongiovanni, G., Fumero, S. & Rossi, L. *J. Chromatogr.* **74**, 255-263 (1972).
8. Charlwood, D.V. & Bell, E.A. *J. Chromatogr.* **135**, 377-384 (1977).
9. Knipfel, J.F., Christensen, D.A. & Owen, B.D. *J. Assoc. Analyt. Chem.* **52**, 981 (1969).

Greg Ogden is at LKB Biochrom Ltd, Cambridge Science Park, Milton Road, Cambridge CB4 4FJ, UK.

Amino acids challenge chromatography

Spurred on by advances in liquid chromatography and derivative chemistry, amino acid analysis systems are beginning to make their mark. Here is a sampling of amino acid specialists along with all-purpose separation supplies.

VARIAN's separation scientists have been busy evaluating the success of 9-fluorenylmethyl chloroformate (Fmoc) derivatization in amino acid analysis. Their latest laboratory update describes optimal conditions for the separation and detection of



Varian reports success with Fmoc analysis.

Fmoc derivatives, as well as the system's potential for automation (Reader Service No. 100). Fmoc, used for many years as an amino protective group in peptide synthesis, also offers several advantages as a precolumn derivatizing agent: its derivatives are stable, results are reproducible, detection limits fall between 100 and 200 fmol and the agent reacts with both primary and secondary amino acids. Extraction is sometimes necessary to remove the Fmoc alcohol by-product of the derivatization reaction. Fmoc alcohol will also show up in chromatograms generated by ultraviolet or fluorescence detection systems. The Varian group performed separations with their own components, including the MicroPak SP C-18 column and the 5500 HPLC system.

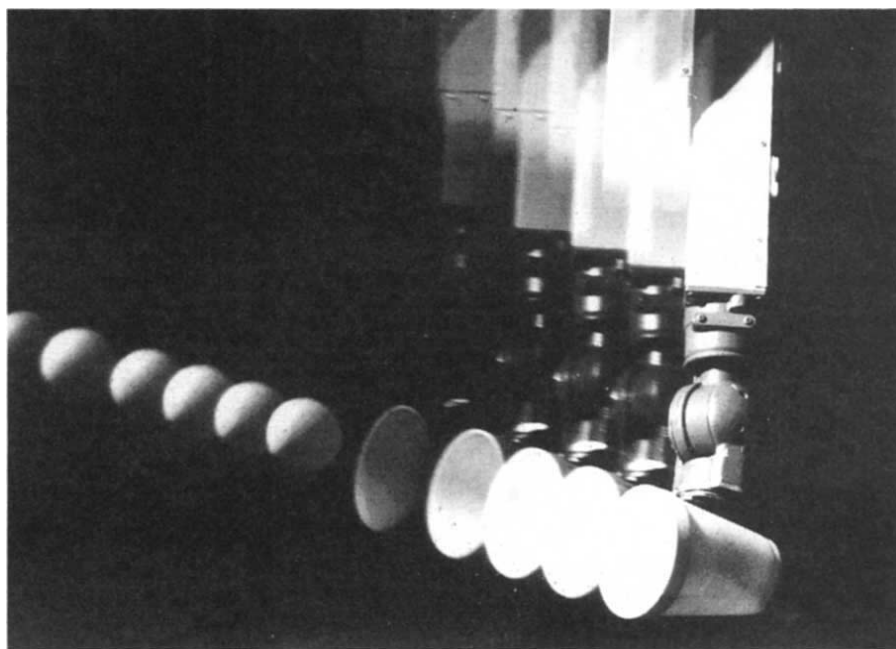
A chiral stationary phase for HPLC can separate optical isomers of alpha-amino acids. Daicel Chemical Industries makes a polymeric material that utilizes ligand exchange chromatography to achieve resolution over a broad range of aliphatic and aromatic DL-amino acids (Reader Service No. 101). Racemates are separated on a 250 mm column with a 4.6 mm internal diameter and fitted with an integral guard column. For complex mixtures, an octadecylsilane column run in series with the Chiralpak W column can improve resolution. A 0.25 mM aqueous copper sulphate mobile phase is the norm, although where required acetonitrile and methanol can be used as retention modifiers. Technicol Ltd distributes the Chiral-

pak column in the United Kingdom at a price of £720 each.

Among the many different types of analytical instruments in Carlo Erba Strumentazione's latest catalogue, their 3A29 amino acid analyser may interest the protein chemist (Reader Service No. 102). The microprocessor-controlled machine will analyse protein or peptide hydrolysates in 50 minutes and physiological fluids in less than 3 hours. Its design is flexible enough to handle the diverse applications of academic research or routine determinations in agriculture, food technology or clinical analysis.

COMPSEQ was developed by the medical biochemistry department of the University of Geneva to put the results of amino acid sequence analysis into a global context. The microcomputer software package contains the protein sequence database of the National Institutes of Health's Protein Identification Resource

(PIR), which it can scan for sequence or composition homologies when presented with an unknown characterization (Reader Service No. 103). COMPSEQ can also compare any two protein sequences for similarity or homology, determine the least redundant nucleotide probe from any given fragment, and predict a sequence's secondary structure or the position of antigenic determinants. The package will run on most advanced 16-bit microcomputers under MS-DOS or PC-DOS operating systems. In the United States, P & S Biochemicals sells the package with a considerable discount for non-profit institutions. COMPSEQ also provides sequencing analysis for nucleic acids. Its menu-driven program will subject digestion fragments to interpretative analysis such as restriction site identification, coding region searches, sequence translation and direct or inverted repeat site location. The EMBL nucleotide sequence data library accompanies the PIR database in COMPSEQ.



Anyone for tennis?

It's not quite ready for Wimbledon, but a robot at AT&T Bell Laboratories knows how to keep its eye on the ball. Robotics systems researchers in Bell's Holmdel, New Jersey site have developed a machine with 'reflexes' that can catch a ping-pong ball on the first bounce. The system uses two pairs of television cameras and high-

speed vision chips to monitor and process the ball's flight. A microprocessor then computes the ball's trajectory and moves the arm into position. The robotic coordination project is part of Bell's efforts to improve robot speed and versatility. Although the ping-pong problem is fairly simple, its technical solutions may help tomorrow's robots to plan and act in more complex and dynamic environments.

Protein peripherals

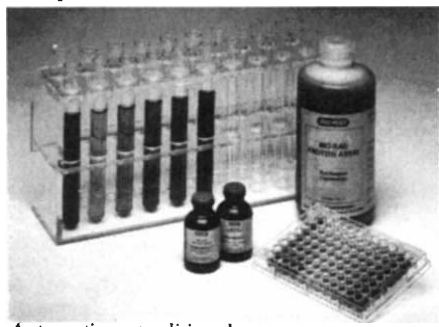
Applied Biosystems is getting away from conventional HPLC for protein and peptide purification. Instead, they have in-



Microbore HPLC minimizes sample dilution.

roduced the model 130A, a separation system that incorporates **microbore liquid chromatography** to yield concentrated products and isolate samples down to the low picomole range (Reader Service No. 104). By generating concentrated samples straight from the purification method, the system circumvents standard concentration techniques that can degrade or destroy samples. Its configuration is well-suited to the micropreparative separation preceding protein or peptide sequencing. Applied Biosystems put the 130A together with proteins in mind, so the \$36,000 (US) system contains column supports selected for optimal protein and peptide separative properties.

Bio-Rad's **protein assay** is just a step away from automation. The same decade-old method, scaled down to fit microtitration plates, can increase the number of



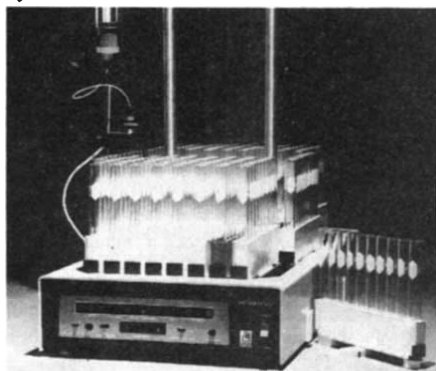
Automating a traditional assay.

assays processed and conserve the amounts of sample used. The plates can then be linked with Bio-Rad's protein assay reader, a mini-spectrophotometer that can make the absorbance measurements for a 96-well plate in less than one minute (Reader Service No. 105). The reader then automatically calculates sample concentration from a standard curve. For less than \$3,000 (US), Bio-Rad's instrument comes with a 595 nm interference filter, a built-in printer and an LED display.

Assistants in analysis

A new **fraction collector** from ISCO can be adapted to collect from up to 20 sources with collecting modes in minutes, tenths of minutes, measured volume or counted

drops in units from 1 to 999 (Reader Service No. 106). The Retriever IV holds racks for either 133 28 mm scintillation vials, 133 25 mm test tubes, 190 10 to 18 mm test tubes or 266 13 mm rimless tubes. Because it is internally heated, the \$1,695 (US) instrument can weather coldroom use down to 0°C. The Retriever will also withstand 100 per cent humidity. It offers void volume diversion, delay timing to synchronize tube contents with chart re-



ISCO's collector wards off coldroom condensation.

corder tracings, LED readouts showing the number of tubes filled and current time/drop/volume count and automatic shutoff.

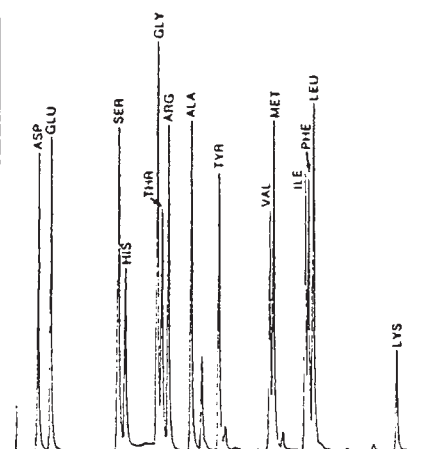
Amino acid derivatization with orthophthalaldehyde (OPA) has several advantages and one disadvantage that recommend it to automation. OPA reacts rapidly in aqueous conditions at ambient temperatures, and it is not necessary to extract the parent compound from the reaction mixture because it is not fluorescent. These characteristics simplify the automation procedures and make OPA an attractive precolumn derivatizing agent.

Precision of precolumn OPA derivatization of amino acids using automix in the Model 9090 LC Auto-Sampler

Amino acid (in order of elution)	% RSD (retention time)	% RSD (peak area)*
Aspartic acid	1.2	2.2
Glutamic acid	1.5	1.0
Serine	1.2	2.1
Histidine	1.0	2.5
Glycine	0.8	1.6
Threonine	0.6	2.3
Arginine	0.7	2.3
Alanine	0.6	1.9
Tyrosine	0.7	1.8
Valine	0.4	1.7
Methionine	0.4	2.1
Isoleucine	0.2	1.2
Phenylalanine	0.2	1.3
Leucine	0.1	1.8
Lysine	0.2	16.8

*Seven separate derivatizations

But OPA derivatives, especially that of lysine, are quite unstable. According to a recent report from Varian, however, this feature is not necessarily a drawback as long as reactions are performed exactly the same way each time — as they are in automated preparation procedures (Reader Service No. 107). Varian reports



Varian peaks with 125 picomoles.

highly reproducible results using their 9090 autosampler (see Table), which can automatically mix a sample and reagent, pause for a prescribed period of time, then inject the sample. Varian's instrument, however, cannot overcome one other major disadvantage to OPA use: it will not form fluorescent derivatives of secondary amino acids.

CHROMATOCHART-PC is Heyden Datasystem's new **chromatography automation software** that can support up to



Heyden's new look for chromatography.

four channels of simultaneous data acquisition and control for GCs, HPLCs and so on (Reader Service No. 108). The menu-driven program enables chromatograms to be displayed on IBM XT or AT monitors, or to be charted on IBMN, Epson or Kipp & Zonen printers. **CHROMATOCHART-PC** will review, edit and store baseline information, and re-analyse data stored in the memory or on a disk, removing the necessity of repeating runs. The program's integration routines measure the area under each peak, determine retention times and calculate peak heights, widths at half height and symmetries. These figures can then be compared to internal or external standards. The £1,350 (UK) software also allows data acquisition and analysis methods to be stored on the disk.

Combining super minicomputers and two new products, Beckman has devised a



Beckman takes LC from digital to analogue.

laboratory automation system for chromatography that starts at less than \$40,000 (US). Beckman's automation scheme consists of their MK5 Digimetry instrument interface coupler and PeakPro software (Reader Service No. 109). The coupler joins chromatograph to computer with analog-to-digital conversion occurring as close to the detector as possible, thereby minimizing noise in the signal. An autoranging feature equates the full scale input range of the coupler with the detector's output signal and, according to Beckman, reduces non-linearity errors in the conversion to less than 0.02 per cent. PeakPro software will process chromatograms with up to 1,000 peaks. Beckman's algorithms identify peaks and assign base-lines automatically, whilst interaction with a mouse-driven cursor is possible to display, examine or construct methods and analyse data.

Packing it in

Buying HPLC equipment is only the beginning: Jaytee Electronic Services provide **on- and off-site servicing** and repairs in southern England for Waters and Altex HPLC equipment (Reader Service No. 110). Jaytee claims competitive rates, sometimes less than half those charged by

the manufacturers themselves. Spare parts for pumps and detectors are also part of Jaytee's business. Sapphire plungers, ball and seat check valves, gaskets, seals and ultraviolet source lamps are sold with full guarantees.

Reusable stainless steel preparative HPLC columns from Separations Technology now come in a range of sizes under the ST 2000 label (Reader Service No. 111). Internal diameters of the columns range from 1 to 10 inches; the ST/LAB 800A system can handle ID's up to 4 inches, ST/LAB 2000 will accommodate sizes up to 6 inches and the ST/PROCESS 200 utilizes columns with ID's from 4 to 10 inches. All the ST 2000 columns are also compatible with Waters 500 and other LC components, and can benefit from the protection of guard columns. Separations Technology's columns, distributed in the United Kingdom by Technicol, cost from £550 to several thousand pounds depending on size. Packing materials can contain particles as small as 10 μm , and sample loading can be increased by joining columns in a series.

Whatman has added four new **ion exchange packings** to their Partisphere



More choices in Partisphere packings.

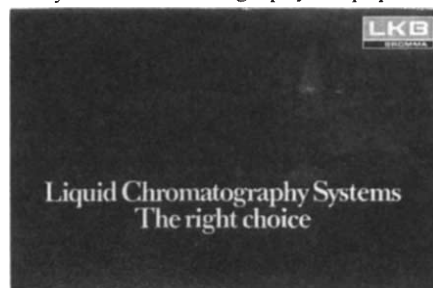
HPLC cartridge systems (Reader Service No. 112). The new package should be particularly advantageous for the analysis of amino acids, nucleotides, polyamines and drugs. At the heart of Whatman's Partisphere system is an integral void sealing designed to give a long cartridge life. Complete systems (cartridges and end fittings) start at £143 (UK). Pre-packed cartridge kits are also available with silica, reversed phase, anion or cation exchange and propyl amino cyano media.

Resin-based columns in hydrogen, calcium, lead, chloride, sodium, silver, hydroxide and neutral resin forms are now available from HPLC Technology (Reader Service No. 113). These diverse Rezex columns can be used in the analysis of sugars, carbohydrates, organic acids and alcohols, as well as amino acids. HPLC Technology presently sells these columns from £265 (UK), with full money-back guarantees.

Chromatography in print

A 16-page guidebook from LKB helps the

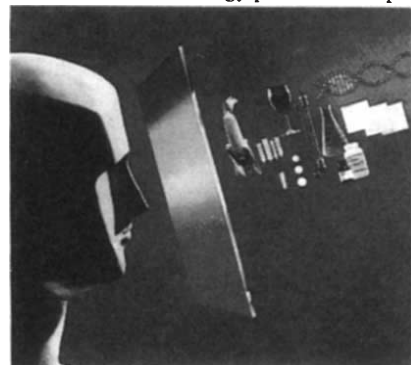
chromatographer navigate the dizzying array of chromatography equipment



LKB has hints for LC selections.

available from modern industry (Reader Service No. 114). *Liquid Chromatography Systems: The Right Choice* has colour illustrations and clearly divided sections to steer researchers through choices in columns, core systems, system peripherals, detectors, output devices and fraction collectors. LKB's brochure outlines the questions a chromatographer should ask and have answered to assemble a customized collection of components.

J.T. Baker's new chemicals catalogue could serve as a handy laboratory reference even if the chemistry closet is full (Reader Service No. 115). *Products for Chromatography* describes Baker's line of materials for sample preparation with solid phase extraction, analytical and preparative HPLC columns and packings, TLC plates and sorbents, high purity solvents and biotechnology products for pro-



A brace of chromatography chemicals.

tein characterization and purification. Applications and guidelines come along with the 96-page guide.

Outstripping J.T. Baker by 40 pages, Chrompack has a *Guide to Chromatography* that features over 100 of their latest wares, along with scientific and applications notes (Reader Service No. 116). Chrompack says their tomb incorporates a format that makes finding a particular product quick and easy. In addition to the 100 new products, the company's entire range of HPLC and GC columns and accessories are also represented. □

These notes are compiled by Karen Wright from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal.

ADVERTISEMENTS

SINGER INSTRUMENTS
MICROMANIPULATORS & ACCESSORIES

**YEAST?
TETRAD ANALYSIS?**



Singer MK III Micromanipulator
and Precision Elevator.

Write direct to:
Singer Instrument Co. Ltd.
Treborough Lodge
Roadwater, Watchet
Somerset TA23 0QL
England. Tel. (0984) 40226
Telex: 337492 Comcab G